

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2000, 18:38:42 : Search time 6423.57 Seconds  
(without alignments)  
2513.359 Million cell updates/sec

Title: US-09-090-672B-1  
Perfect score: 4276  
Sequence: 1 TTCTACCGTTTTTCCCTGC.....ATCAGAAAAA 4276

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*

- 44: gb\_est25:\*  
45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	386	9.0	506	45	AI392674 tg47c02.x

High quality sequence stop: 444.		Location/Qualifiers		1..506		/organism="Homo sapiens"		/db_xref="taxon:9606"		/clone="IMAGE:2111906"		/clone_lib="Soares_NFL_T-GBC_S1"		/lab_host="DH10B"		/note="Organ: pooled; Vector: pT7M3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19w, testis NHT, and B-cell NCI-CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaudo."		BASE COUNT		185 a 75 c 82 g 164 t		ORIGIN		Query Match		9.0%; Score 386; DB 45; Length 506;		Best Local Similarity 99.8%; Pred. No. 1.2e-13;		Matches 506; Conservative		Mismatches 0; Indels 1; Gaps 1;																																																																																							
QY 3762		ACAATATGAACCTCTTTTTCGTATGCCATCGGGTGCATGGAAGTTTATCTCTGTT		3821		DB 506		ACAATATGAACCTCTTTTTCGTATGCCATCGGGTGCATGGAAGTTTATCTCTGTT		447		QY 3822		TTGCTGGAACCAAGAGATCCAACTTCTGCAACATTTCTTAGAGGAGAGAGAGAAA		3881		DB 446		TTGCTGGAACCAAGAGATCCAACTTCTGCAACATTTCTTAGAGGAGAGAGAGAAA		387		QY 3882		TATTAAG		3941		DB 386		TATTAAG		327		QY 3942		ATATAACATATAAGATACATTTTATTAATAACCATGCAACAATAACACTATCGGCTA		4001		DB 326		ATATAACATATAAGATACATTTTATTAATAACCATGCAACAATAACACTATCGGCTA		267		QY 4002		TCTGACAGTTTTCCTCCAGGAGAGTCTTTTGCCTTTTCTTTTCTTTTCTTTTCTTTT		4061		DB 266		TCTGACAGTTTTCCTCCAGGAGAGTCTTTTGCCTTTTCTTTTCTTTTCTTTTCTTTT		207		QY 4062		CATCTTTTCTCTCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAG		4121		DB 206		CATCTTTTCTCTCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAG		148		QY 4122		GAAGTTAAACAATTTTAAATCGAAGAGCATGTTAGAGCAACAATAACACTATCGGCTA		4181		DB 147		GAAGTTAAACAATTTTAAATCGAAGAGCATGTTAGAGCAACAATAACACTATCGGCTA		88		QY 4182		GAGCAGCATTAATAATTTTCAGGGTTTTCAGGCTGCAACATAATTCATTATCCCTCA		4241		DB 87		GAGCAGCATTAATAATTTTCAGGGTTTTCAGGCTGCAACATAATTCATTATCCCTCA		28		QY 4242		AAAACTTACCACCACATCAGAAAAAAA		4268		DB 27		AAAACTTACCACCACATCAGAAAAAAA		1		RESULT 2		H71225		398 bp		mrna		EST		26-OCT-1995	
LOCUS		H71225		398 bp		mrna		EST		26-OCT-1995		DEFINITION		Ys12e09.r1 Soares fetal liver spleen lNfLS Homo sapiens cDNA clone IMAGE:214600 5' similar to SP:040H1.1 C500109 OVARIAN PROTEIN ; , mRNA sequence.		H71225		ACCESSION		H71225		VERSION		H71225.1		GI:1043041																																																																																													















AI093858/c  
 LOCUS AI093858 365 bp mRNA EST 18-AUG-1998  
 DEFINITION ga30d05.s1 Soares\_NhHMPu\_S1 Homo sapiens cDNA clone IMAGE:1688265  
 3', mRNA sequence.  
 ACCESSION AI093858  
 VERSION AI093858  
 KEYWORDS EST.  
 SOURCE AI093858.1 GI:3432834  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 365)  
 AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2152930.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: 40m13 fwd. BT from Amersham  
 High quality sequence stop: 339.  
 Location/Qualifiers  
 1..365  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1688265"  
 /clone\_lib="Soares\_NhHMPu\_S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and  
 pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not  
 Site 2: Eco RI; Equal amounts of plasmid DNA from three  
 normalized libraries (melanocyte 2NBHM, pregnant uterus  
 NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
 were made in vitro. Following HAP purification, this DNA  
 was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from pools of  
 5,000 clones made from the same 3 libraries. The pools  
 consisted of 1.M.A.G.E. clones 260232-265223,  
 340488-343479, and 484488-489479."  
 340488-343479, 60 g 114 t  
 144 a 47 c  
 BASE COUNT 144 a 47 c 60 g 114 t  
 ORIGIN  
 Query Match 4.3%; Score 186; DB 42; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-60;  
 Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4074 TCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAGTTAAACAT 4133  
 Db 186 TCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAGTTAAACAT 127  
 QY 4134 TTTTAATGGAAGAGCAGTTAGAGCAACAATGATTAAGCAAGCTGAGCAGCATTTAT 4193  
 Db 126 TTTTAAGGAAGAGCAGTTAGAGCAACAATGATTAAGCAAGCTGAGCAGCATTTAT 67  
 QY 4194 ARTTAATTTTCAGGTTTGGAGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 4253  
 Db 66 ARTTAATTTTCAGGTTTGGAGCTGAACATAATTTTCATTATCCCTCAAAAAGTTACCAC 7  
 QY 4254 CACATC 4259  
 Db 6 CACATC 1  
 RESULT 14  
 AI0962924/c  
 LOCUS AI0962924 433 bp mRNA EST 20-AUG-1999  
 DEFINITION wt24h06.x1 NCI-CCGAP\_U01 Homo sapiens cDNA clone IMAGE:2506443 3',

mRNA sequence.  
 ACCESSION AI0962924  
 VERSION AI0962924.1 GI:5755637  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 433)  
 AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1134595.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 362.  
 Location/Qualifiers  
 1..433  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2506443"  
 /clone\_lib="NCI-CCGAP\_U01"  
 /tissue\_type="well-differentiated endometrial  
 adenocarcinoma, 7 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;  
 Site 2: NotI; Cloned unidirectionally. Primer: oligo 47.  
 Average insert size 1.75 kb. Life Technologies catalog #:  
 11538-014"  
 153 a 59 c 68 g 151 t 2 others  
 BASE COUNT 153 a 59 c 68 g 151 t 2 others  
 ORIGIN  
 Query Match 4.1%; Score 175; DB 63; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-56;  
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4074 TCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAGTTAAACAT 4133  
 Db 205 TCTCTCTTTTTCATCCCTTTTAAATTTTAAACAGCAATGGAGGAGTTAAACAT 146  
 QY 4134 TTTTAATGGAAGAGCAGTTAGAGCAACAATGATTAAGCAAGCTGAGCAGCATTTAT 4193  
 Db 145 TTTTAATGGAAGAGCAGTTAGAGCAACAATGATTAAGCAAGCTGAGCAGCATTTAT 86  
 QY 4194 AATTAATTTTCAGGTTTGGAGCTGAACATAATTTTCATTATCCCTCAAAAAGTT 4248  
 Db 85 AATTAATTTTCAGGTTTGGAGCTGAACATAATTTTCATTATCCCTCAAAAAGTT 31  
 RESULT 15  
 AA381125/c  
 LOCUS AA381125 156 bp mRNA EST 21-APR-1997  
 DEFINITION EST94220 Activated T-cells I Homo sapiens cDNA 3' end, mRNA  
 sequence.  
 ACCESSION AA381125  
 VERSION AA381125.1 GI:2033495  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;





AT1762436/c 568 bp mRNA EST 24-JUN-1999  
 LOCUS w957a05.x1 Soares\_NSFF8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2369168.3, mRNA sequence.  
 ACCESSION AT1762436.1 GI:5178103  
 VERSION EST.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 568)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3187416.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 483.  
 FEATURES  
 Location/Qualifiers  
 1..568  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2369168"  
 /clone\_lib="Soares\_NSFF8\_9W\_OT\_PA\_P\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from five normalized  
 libraries were mixed, and ss circles were made in vitro.  
 Following HAP purification, this DNA was used as tracer in  
 a subtractive hybridization reaction. The driver was  
 PCR-amplified cDNAs from pools of 5,000 clones made from  
 the same 5 libraries. The pools consisted of the following  
 libraries and clones: Soares NBHSF pool 1:  
 303984-310935, 323208-325895 Soares NB2HF pool 1:  
 145032-147939, 147720-148103, 148872-149255, 15002 -  
 150407, 151176-152327 Soares NB2HF-9W pool 1:  
 758280-760583, 772104-774407 Soares NBHPA pool 1:  
 304776-306311, 320136-322823, 326280-326663 Soares NBHOT  
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 BASE COUNT 230 a 79 c 88 g 170 t 1 others  
 ORIGIN  
 Query Match 2.2%; Score 93; DB 51; Length 568;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-25;  
 Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 205 AGATGAGATCAGCTAGTTTCGTCGCTTGCGCCCTCTGATTTGGATTGGCCTCATAA 264  
 Db 409 AGATGAGATCAGCTAGTTTCGTCGCTTGCGCCCTCTGATTTGGATTGGCCTCATAA 350  
 QY 265 AGCTGAGACCAATCTATTTCCTCCTAAAGG 297  
 Db 349 AGCTGAGACCAATCTATTTCCTCCTAAAGG 317  
 RESULT 20  
 B55278/c 639 bp DNA GSS 20-JUN-1998  
 LOCUS CIT-HSP-386K14.TR CIT-HSP Homo sapiens genomic clone 386K14,  
 DEFINITION genomic survey sequence.  
 ACCESSION B55278  
 VERSION B55278.1 GI:2609612  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 764)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 639)  
 AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
 Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other GSSs: CIT-HSP-386K14.TF  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdamams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 clones are available from  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: M13 Reverse  
 Class: BAC ends.  
 FEATURES  
 Location/Qualifiers  
 1..639  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5379185"  
 /db\_xref="taxon:9606"  
 /clone="386K14"  
 /clone\_lib="CIT-HSP"  
 /sex="Male"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII"  
 BASE COUNT 197 a 141 c 159 g 142 t  
 ORIGIN  
 Query Match 1.2%; Score 53; DB 81; Length 639;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-10;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1894 CACCTCTCCCTCCGGGTTCAAGTGAATTCCTCTAGCTTAGCTCCCGAGTAG 1946  
 Db 361 CACCTCTCCCTCCGGGTTCAAGTGAATTCCTCTAGCTTAGCTCCCGAGTAG 309  
 RESULT 21  
 AQ744981/c 764 bp DNA GSS 16-JUL-1999  
 LOCUS HS\_5501A2.C03.SP6.RPCI-11 Human Male BAC Library Homo sapiens  
 DEFINITION genomic clone Plate-1077 Col-6 Row-E, genomic survey sequence.  
 ACCESSION AQ744981  
 VERSION AQ744981.1 GI:5522591  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 764)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and  
 scanning the human genome  
 JOURNAL Proc.Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887



Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>  
 Plate: 1077 row: E column: 6  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence stop: 764.

Location/Qualifiers

1..764  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=1077 Col=6 Row=E"  
 /clone\_lib="RPCI-11 Human Male BAC Library"  
 /sex="male"  
 /note="Vector: pBACe3.6; Genomic sequence of BAC ends"

FEATURES  
 source

BASE COUNT 233 a 138 c 164 g 228 t 1 others  
 ORIGIN

Query Match 1..18; Score 47; DB 87; Length 764;

Best Local Similarity 100.0%; Pred. No. 5.7e-08; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1940  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 422 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 376

RESULT 22

LOCUS AQ061635 261 bp DNA GSS 31-JUL-1998  
 DEFINITION CIT-HSP-234805 RF CIT-HSP Homo sapiens genomic clone 234805,  
 genomic survey sequence.

ACCESSION AQ061635

VERSION AQ061635.1 GI:3363547

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 261)

Ahrens,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,

Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,

Simon,M., and Venter,J.C. Sequence Database for Sequence-Ready Map

Building (1998)

Unpublished (1998)

Other\_GSSs: CIT-HSP-234805.TR

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1..261

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="234805"

/clone\_lib="CIT-HSP"

/sex="Male"

/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:

FEATURES  
 source

BASE COUNT 42 a 62 c 66 g 91 t

ORIGIN

Query Match 1..18; Score 46; DB 90; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1939  
 ||||||||||||||||||||||||||||||||||||||||||||||||

Db 107 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 152

RESULT 23

LOCUS AA350648

DEFINITION EST58017 Infant brain Homo sapiens cDNA 3' and similar to EST

containing Alu repeat, mRNA sequence.

ACCESSION AA350648

VERSION AA350648.1 GI:2002965

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 269)

Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.

3,400 expressed sequence tags identify diversity of transcripts

from human brain

Nature Genet. 4, 256-267 (1993)

JOURNAL 9364420

MEDLINE

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1407373.

Other\_ESTs: PHC83368

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MO 20850 USA

Tel: 3018699036

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13-21.

Location/Qualifiers

1..269

/organism="Homo sapiens"

/db\_xref="AYCC (Inhost):151254"

/db\_xref="taxon:9606"

/clone\_lib="Infant brain"

/sex="female"

/dev\_stage="infant"

/note="Organ: brain; Vector: lafmid BA; Site\_1: HindIII;

Site\_2: NotI"

BASE COUNT 54 a 78 c 66 g 71 t

ORIGIN

Query Match 1..18; Score 45; DB 32; Length 269;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 1939  
 ||||||||||||||||||||||||||||||||||||||||||||||||

Db 61 CAACCTCTGCTCCCGGGTTCAAGTGATTCTCTCGCTTAGCCTCC 106

RESULT 24

LOCUS AA955031

DEFINITION oq66h02.s1 NCT\_CGAP\_K166 Homo sapiens cDNA clone IMAGE:1591347 3'

similar to contains Alu repetitive element;contains element MER22

EST 17-MAR-1999

oq66h02.s1 NCT\_CGAP\_K166 Homo sapiens cDNA clone IMAGE:1591347 3'

similar to contains Alu repetitive element;contains element MER22

repetitive element ;, mRNA sequence.

AA955031  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AA955031.1 GI:3118726  
EST.  
Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 343)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19 1998 this sequence version replaced gi:2151892.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 495-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Stratagene, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Insert Length: 2046 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 320.

FEATURES  
source  
Location/Qualifiers  
1..343  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1591347"  
/clone\_lib="NCI-CGAP\_kid6"  
/sex="mixed"  
/tissue\_type="kidney tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: kidney; Vector: Bluescript SK-; Site: 1;  
Scori: Site 2; XhoI: Cloned unidirectionally. Primer:  
oligo dt. Pooled kidney tumors. 5' adaptor sequence: 5'  
GAATTCGGCAGCAG 3' 3' adaptor sequence: 5'  
CTCAGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."  
BASE COUNT 59 a 103 c 88 g 93 t  
ORIGIN

Query Match 1.1%; Score 46; DB 40; Length 343;  
Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCTCCCGGGTTCAAGTGATTCCTGCTTACGCTCC 1939  
|||||  
Db 70 CAACCTCTGCTCCCGGGTTCAAGTGATTCCTGCTTACGCTCC 115  
sequence.

RESULT 25  
LOCUS  
DEFINITION

AQ608513 434 bp DNA GSS 10-JUN-1999  
HS\_2124\_B2\_B09\_77C CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate-2124 Col-18 Row-D, genomic survey  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AQ608513  
AQ608513.1 GI:5068507  
GSS.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 434)  
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., and  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, W.D., and  
Hood, L.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 2124 row: D column: 18  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 434.  
Location/Qualifiers  
1..434  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-2124 Col-18 Row=D"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 124 a 111 c 104 g 93 t 2 others  
ORIGIN

Query Match 1.1%; Score 46; DB 105; Length 434;  
Best Local Similarity 100.0%; Pred. No. 1.7e-07;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1886 GGTCACTCAACCTCTGCTCCCGGGTTCAAGTGATTCCTGCTCC 1931  
|||||  
Db 252 GGTCACTCAACCTCTGCTCCCGGGTTCAAGTGATTCCTGCTCC 207  
sequence.

RESULT 26  
LOCUS  
DEFINITION

R50700 426 bp mRNA EST 18-MAY-1995  
YJ60912.r1 Soares breast 2NDHBst Homo sapiens cDNA clone  
IMAGE:153190 5', similar to gb:X59434 THIOSULFATE SULFURTRANSFERASE  
(HUMAN); contains Alu repetitive element; contains MSRI repetitive  
element ;, mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

R50700  
R50700.1 GI:812602  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 426)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Travaaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and  
Wilson, R.  
The WashU-Merck EST Project  
Contact: wilson RK  
Unpublished (1995)  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 926  
High quality sequence stops: 203 Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 926 Std Error: 0.00  
Seq primer: M13Rp1



```

RESULT 29
AA654395      180 bp      mRNA      EST      23-DEC-1997
LOCUS
DEFINITION nt03e09.sl NCI_CGAP_Lym3 Homo sapiens cDNA clone IMAGE:1192072 3'
similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION AA654395
VERSION AA654395.1 GI:2590549
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520479.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
WARNING: There is evidence that suggests that the 384-well parent
plate of this clone contains both human and mouse derived clones.
Thus, the origin of this clone is uncertain. This caution should be
kept in mind should you use this clone.

Tissue Procurement: Mark Raffeld, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1..180
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1192072"
/clone_lib="NCI_CGAP_Lym3"
/tissue_type="lymphoma"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: lymph node; Vector: Bluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Pooled lymphomas. 5' adaptor sequence: 5'
GAATTCGGCAGG 3' 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 0.9 kb."
BASE COUNT 32 a 52 c 46 g 50 t
ORIGIN
CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 1945
|||||
|||||

Query Match 1.0%; Score 43; DB 36; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.3e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1903 CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 1945
|||||
|||||

Db 70 CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 112
|||||
|||||

RESULT 30
AI926033      383 bp      mRNA      EST      02-SEP-1999
LOCUS
DEFINITION w01b09.x1 NCI_CGAP_Gas4 Homo sapiens cDNA clone IMAGE:2457881 3'
similar to contains Alu repetitive element;contains element MER9
repetitive element.; mRNA sequence.
ACCESSION AI926033
VERSION AI926033.1 GI:5661997
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS Hillier.L., Lennon.G., Becker.M., Bonaldo.M.F., Chiapelli.B.,
Chisoe.S., Dietrich.N., Dubuque.T., Favello.A., Gish.W.,
Hawkins.M., Hultman.M., Kucaba.T., Lacy.M., Le.M., Le.N.,
Mardis.E., Moore.B., Morris.M., Parsons.J., Prange.C., Rifkin.L.,
Rohlfing.T., Schellenberg.K., Soares.M.B., Tan.F., Thierry-Mieg.J.,
Trevisan.E., Underwood.K., Wohlmann.P., Waterston.R., Wilson.R.,
and Marra.M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
TITLE
JOURNAL

```

```

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 383)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137633.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 375.
Location/Qualifiers
1..383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2457881"
/clone_lib="NCI_CGAP_Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/notes="Organ: stomach; Vector: pCMV-SPOK6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"
BASE COUNT 85 a 103 c 89 g 105 t 1 others
ORIGIN
CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 1945
|||||
|||||

Query Match 1.0%; Score 43; DB 62; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.4e-06;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1903 CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 1945
|||||
|||||

Db 71 CTCCCGGGTCAAGTGATTCTCTCGCTTAGCCTCCCGAGTA 113
|||||
|||||

RESULT 31
AA121819/c      391 bp      mRNA      EST      02-FEB-1997
LOCUS
DEFINITION zn95o09.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone
IMAGE:565984 5' similar to contains Alu repetitive element;contains
element MER22 repetitive element.; mRNA sequence.
ACCESSION AA121819
VERSION AA121819.1 GI:1679469
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS Hillier.L., Lennon.G., Becker.M., Bonaldo.M.F., Chiapelli.B.,
Chisoe.S., Dietrich.N., Dubuque.T., Favello.A., Gish.W.,
Hawkins.M., Hultman.M., Kucaba.T., Lacy.M., Le.M., Le.N.,
Mardis.E., Moore.B., Morris.M., Parsons.J., Prange.C., Rifkin.L.,
Rohlfing.T., Schellenberg.K., Soares.M.B., Tan.F., Thierry-Mieg.J.,
Trevisan.E., Underwood.K., Wohlmann.P., Waterston.R., Wilson.R.,
and Marra.M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
TITLE
JOURNAL

```

MEDLINE  
COMMENT  
97044478  
On May 5, 1995 this sequence version replaced gi:798452.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL : contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1369 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 240.  
Location/Qualifiers  
1..391  
/organism="Homo sapiens"  
/db\_xref="GDB:4591920"  
/db\_xref="taxon:9606"  
/clone="IMAGE:565984"  
/clone\_lib="Stratagene fetal retina 937202"  
/sex="mixed"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled  
retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR  
Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'  
adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 103 a 92 c 108 g 78 t 10 others  
ORIGIN

Query Match 1.0%; Score 43; DB 28; Length 391;  
Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1904 CTCGGGTTCAAGTGATTCTCTGCTTAGCTCCGAGTAG 1946  
|||||  
Db 152 CTCGGGTTCAAGTGATTCTCTGCTTAGCTCCGAGTAG 110  
|||||

RESULT 32  
A0515717  
LOCUS A0515717 406 bp DNA GSS 05-MAY-1999  
DEFINITION HS\_5235\_A2\_E01\_T7A RPI-11 Human Male BAC Library Homo sapiens  
genomic clone Plate-811 Col=2 Row=I, genomic survey sequence.  
ACCESSION A0515717  
VERSION A0515717.1 GI:4747975  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 406)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:  
http://www.htsc.washington.edu

Plate: 811 row: I column: 2  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 406.  
Location/Qualifiers  
1..406  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate=811 Col=2 Row=I"  
/clone\_lib="RPI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBACE3.6; Genomic sequence of BAC ends"  
BASE COUNT 77 a 92 c 85 g 148 t 4 others  
ORIGIN

Query Match 1.0%; Score 43; DB 104; Length 406;  
Best Local Similarity 100.0%; Pred. No. 2.4e-06;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1898 CTCGCTCCCGGTTCAAGTGATTCTCTGCTTAGCTCC 1940  
|||||  
Db 348 CTCGCTCCCGGTTCAAGTGATTCTCTGCTTAGCTCC 390  
|||||

RESULT 33  
A047553  
LOCUS A047553 470 bp mRNA EST 29-SEP-1999  
DEFINITION DKF2p58E0121\_s1 586 (synonym: hutel) Homo sapiens cDNA clone  
DKF2p58E0121, mRNA sequence.  
ACCESSION A047553  
VERSION A047553.1 GI:4728549  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 470)  
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and  
Wiemann,S.  
TITLE EST (Ottenwaelder, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189405.  
Contact: Ottenwaelder B  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 3' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by Medigenomix within the cDNA sequencing consortium of  
the German Genome Project.  
rl sequence also available.  
This clone is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..470  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp58E0121"  
/clone\_lib="586 (synonym: hutel)"  
/tissue\_type="uterus"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="vector: pSport1; Site\_1: NotI; Site\_2: SalI/MluI"  
BASE COUNT 125 a 108 c 94 g 143 t  
ORIGIN

Query Match 1.0%; Score 43; DB 49; Length 470;  
Best Local Similarity 100.0%; Pred. No. 2.3e-06;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1903 CCTCCCGGTTCAAGTGATTCTCTCGCTTAGCTCCCGAGTA 1945
      |||||||
Db 86 CCTCCCGGTTCAAGTGATTCTCTCGCTTAGCTCCCGAGTA 128

RESULT 34
AQ394316/c
LOCUS      AQ394316      527 bp      DNA      GSS      06-MAR-1999
DEFINITION CITBI-E1-2546J9.TF CITBI-E1 Homo sapiens genomic clone 2546J9,
            genomic survey sequence.
ACCESSION  AQ394316
VERSION     AQ394316.1 GI:4365339
KEYWORDS   GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 527)
AUTHORS     Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuwa,H., Simon,M. and
            Venter,J.C.
TITLE       Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2546J9.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.

FEATURES             source
    source
        1..527
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="2546J9"
            /clone_lib="CITBI-E1"
            /sex="male"
            /cell_type="sperm"
            /note="vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
            Caltech Human BAC Library D"
BASE COUNT  133 a 107 c 135 g 152 t
ORIGIN

FEATURES             source
    source
        1..527
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="2546J9"
            /clone_lib="CITBI-E1"
            /sex="male"
            /cell_type="sperm"
            /note="vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
            Caltech Human BAC Library D"
BASE COUNT  133 a 107 c 135 g 152 t
ORIGIN

Query Match      1.0%; Score 42; DB 102; Length 527;
Best Local Similarity 100.0%; Pred. No. 5.2e-06;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1894 CAACCTCTGCCTCCCGGTTCAAGTGATTCTCTCGCTTAGC 1935
      |||||||
Db 503 CAACCTCTGCCTCCCGGTTCAAGTGATTCTCTCGCTTAGC 462

RESULT 35
AQ205998/c
LOCUS      AQ205998      389 bp      DNA      GSS      17-SEP-1998
DEFINITION HS_3237_B2_A06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3237 Col=12 Row=B, genomic survey
            sequence.
ACCESSION  AQ205998
VERSION     AQ205998.1 GI:3616568
KEYWORDS   GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 389)

```

```

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3237 row: B column: 12
Class: BAC ends
High quality sequence stop: 389.
Location/Qualifiers
    1..389
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="plate=3237 Col=12 Row=B"
        /clone_lib="CIT Approved Human Genomic Sperm Library D"
        /sex="male"
        /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
        E-Coli DH10B"
BASE COUNT  123 a 100 c 87 g 77 t 2 others
ORIGIN

Query Match      1.0%; Score 41; DB 99; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1900 CTGCTCCCGGTTCAAGTGATTCTCTCGCTTAGCTCCCTCCC 1940
      |||||||
Db 136 CTGCTCCCGGTTCAAGTGATTCTCTCGCTTAGCTCCCTCCC 96

RESULT 36
AQ206027/c
LOCUS      AQ206027      400 bp      DNA      GSS      17-SEP-1998
DEFINITION HS_3237_B2_D05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3237 Col=10 Row=H, genomic survey
            sequence.
ACCESSION  AQ206027
VERSION     AQ206027.1 GI:3616597
KEYWORDS   GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 400)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3237 row: H column: 10
Class: BAC ends
High quality sequence stop: 400.
Location/Qualifiers

```

```
source
1. .400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3237 Col=10 Row=H"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 135 a 94 c 89 g 82 t
ORIGIN

Query Match 1.0%; Score 41; DB 99; Length 400;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1900 CTGCCTCCGGGTTCAAGTGATTCTCCGCTTAGCCTCCC 1940
Db 136 CTGCCTCCGGGTTCAAGTGATTCTCCGCTTAGCCTCCC 96

RESULT 37
A0763211/c 513 bp DNA GSS 28-JUL-1999
LOCUS HS_3222_B2_E12_77C CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=3222 Col=24 Row=J, genomic survey
sequence.
ACCESSION A0763211
VERSION A0763211.1 GI:5641327
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 513)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3222 Row: J column: 24
Seq primer: 17
Class: BAC ends
High quality sequence stop: 513.
Location/Qualifiers
1. .513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3222 Col=24 Row=J"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/notes="Organ: sperm; Vector: pBelOBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 166 a 118 c 117 g 110 t 2 others
ORIGIN

Query Match 1.0%; Score 41; DB 87; Length 513;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1900 CTGCCCTCCGGGTTCAAGTGATTCTCCGCTTAGCCTCCC 1940
```

```
Db 154 CTGCCCTCCGGGTTCAAGTGATTCTCCGCTTAGCCTCCC 114

RESULT 38
A0742061/c 693 bp DNA GSS 16-JUL-1999
LOCUS HS_5570_B2_H12_SP6 RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=1146 Col=24 Row=P, genomic survey sequence.
ACCESSION A0742061
VERSION A0742061.1 GI:5519583
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piederdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1146 row: P column: 24
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 693.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1146 Col=24 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 231 a 131 c 160 g 171 t
ORIGIN

Query Match 1.0%; Score 41; DB 82; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGTTCAAGTGATTCTCTCGCTTAG 1934
Db 494 CAACCTCTGCCTCCCGGTTCAAGTGATTCTCTCGCTTAG 454

RESULT 39
A0268072 104 bp DNA GSS 27-APR-1999
LOCUS RPC111-7317.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-7317,
DEFINITION genomic survey sequence.
ACCESSION A0268072
VERSION A0268072.1 GI:3795676
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 104)  
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,  
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.  
Use of Human BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1998)  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@igr.org  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1. .104  
/organism="Homo sapiens"  
/db\_xref="GDB:7527846"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-7317"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"  
20 a 37 c 21 g 26 t

Query Match 0.98; Score 39; DB 100; Length 104;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1902 GCCTCCCGGGTCAAGTGATTCTCTGCTTAGCTAGCTCCC 1940  
|||||  
Db 29 GCCTCCCGGGTCAAGTGATTCTCTGCTTAGCTAGCTCCC 67

RESULT 40  
AA054055  
LOCUS  
DEFINITION  
IMAGE:380196 5' similar to contains Alu repetitive element;; mRNA  
sequence.  
AA054055  
VERSION  
AA054055.1 GI:1544979  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 274)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le.N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Roifling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
On May 5, 1995 this sequence version replaced gi:797877.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 2948 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 230.

FEATURES  
source  
1. .274  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="GDB:1288453"  
/db\_xref="taxon:9606"  
/clone="IMAGE:380196"  
/clone\_lib="Soares retina N2b4HR"  
/sex="male"  
/tissue\_type="retina"  
/dev\_stage="55 year old"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: eye; Vector: pT73B (Pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTITTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). The retinas were obtained from a 55 year old  
Caucasian and total cellular poly(A)+ RNA was extracted 6  
hrs after their removal. The retina RNA was kindly  
provided by Roderick R. McInnes M.D. Ph.D. from the  
University of Toronto. Library constructed by Bento  
Soares and M.Fatima Bonaldo.  
52 a 79 c 67 g 75 t 1 others

BASE COUNT 52 a 79 c 67 g 75 t 1 others  
ORIGIN

Query Match 0.98; Score 39; DB 27; Length 274;  
Best Local Similarity 100.0%; Pred. No. 9.2e-05;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1902 GCCTCCCGGGTCAAGTGATTCTCTGCTTAGCTAGCTCCC 1940  
|||||  
Db 65 GCCTCCCGGGTCAAGTGATTCTCTGCTTAGCTAGCTCCC 103

RESULT 41  
AQ100978/c  
LOCUS  
DEFINITION  
HS\_3061\_A1\_A03\_MR CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate=3061 Col=5 Row=A, genomic survey  
sequence.  
AQ100978  
VERSION  
AQ100978.1 GI:3472007  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 301)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu

TITLE  
JOURNAL  
MEDLINE  
COMMENT



Sequence Tagged Connector  
 Plate: 3061 row: A column: 5  
 Class: BAC ends  
 High quality sequence stop: 301.  
 Location/Qualifiers

## FEATURES

source

1. .301  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate=3061 Col-5 Row-A"  
 /clone\_lib="CII Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT

ORIGIN

81 a 61 c 79 g 80 t

Query Match 0.9%; Score 39; DB 91; Length 301;

Best Local Similarity 100.0%; Pred. No. 8.9e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1902 GCTCCCGGTTCAAGTATCTCCGCTAGCCTCCC 1940

Db 299 GCTCCCGGTTCAAGTATCTCCGCTAGCCTCCC 261

RESULT 42

AA084176

LOCUS

DEFINITION

zn17d10.s1 StrataGene neuroepithelium NT2RAM1 937234 Homo sapiens  
 cDNA clone IMAGE:547699 3' similar to contains Alu repetitive  
 element; , mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

(Ntera-2/c1.D1) precursor cells induced with Retinoic  
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
 Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'

BASE COUNT

ORIGIN

20 a 35 c 32 g 26 t 1 others

Query Match 0.9%; Score 38; DB 28; Length 115;

Best Local Similarity 100.0%; Pred. No. 0.00031; Mismatches 0; Indels 0; Gaps 0;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 1931

Db 35 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 72

RESULT 43

H84675/c

LOCUS

DEFINITION

H84675 132 bp mRNA EST 14-NOV-1995  
 rs56d04.r1 Soares retina N2B4HR Homo sapiens cDNA clone  
 IMAGE:219751 5' similar to contains Alu repetitive element; contains  
 MER22 repetitive element ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

Generation and analysis of 280,000 human expressed sequence tags  
 Genome Res. 6 (9), 807-828 (1996)  
 On May 8, 1995 this sequence version replaced gi:801257.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1392 Std Error: 0.00  
 Seq primer: -40M13 fwd. from Amersham  
 High quality sequence stop: 97.  
 Location/Qualifiers

source

1. .115  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate=3061 Col-5 Row-A"  
 /clone\_lib="CII Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT

ORIGIN

81 a 61 c 79 g 80 t

Query Match 0.9%; Score 39; DB 91; Length 301;

Best Local Similarity 100.0%; Pred. No. 8.9e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1902 GCTCCCGGTTCAAGTATCTCCGCTAGCCTCCC 1940

Db 299 GCTCCCGGTTCAAGTATCTCCGCTAGCCTCCC 261

RESULT 42

AA084176

LOCUS

DEFINITION

zn17d10.s1 StrataGene neuroepithelium NT2RAM1 937234 Homo sapiens  
 cDNA clone IMAGE:547699 3' similar to contains Alu repetitive  
 element; , mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

(Ntera-2/c1.D1) precursor cells induced with Retinoic  
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
 Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'

BASE COUNT

ORIGIN

20 a 35 c 32 g 26 t 1 others

Query Match 0.9%; Score 38; DB 28; Length 115;

Best Local Similarity 100.0%; Pred. No. 0.00031; Mismatches 0; Indels 0; Gaps 0;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 1931

Db 35 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 72

RESULT 43

H84675/c

LOCUS

DEFINITION

H84675 132 bp mRNA EST 14-NOV-1995  
 rs56d04.r1 Soares retina N2B4HR Homo sapiens cDNA clone  
 IMAGE:219751 5' similar to contains Alu repetitive element; contains  
 MER22 repetitive element ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

(Ntera-2/c1.D1) precursor cells induced with Retinoic  
 Acid for 1 week, followed by 3 weeks in mitotic inhibitors  
 (Replate #2). Average insert size: 1.1 kb; Uni-ZAP XR  
 Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3'  
 adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'

BASE COUNT

ORIGIN

20 a 35 c 32 g 26 t 1 others

Query Match 0.9%; Score 38; DB 28; Length 115;

Best Local Similarity 100.0%; Pred. No. 0.00031; Mismatches 0; Indels 0; Gaps 0;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 1931

Db 35 CAACCTCTGCCTCCCGGTTCAAGTATCTCTCTGCT 72

RESULT 43

H84675/c

LOCUS

DEFINITION

H84675 132 bp mRNA EST 14-NOV-1995  
 rs56d04.r1 Soares retina N2B4HR Homo sapiens cDNA clone  
 IMAGE:219751 5' similar to contains Alu repetitive element; contains  
 MER22 repetitive element ;, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

TITLE

hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

29 a 34 c 45 g 24 t

BASE COUNT 25 a 54 c 37 g 34 t 1 others

ORIGIN

Query Match 0.9%; Score 38; DB 24; Length 132;  
Best Local Similarity 100.0%; Pred. NO. 0.00029;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 1931  
|||||  
Db 74 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 37

RESULT 44  
AA397419 157 bp mRNA EST 13-AUG-1997  
LOCUS nc64f11.s1 NCI\_CGAP\_Prl Homo sapiens cDNA clone IMAGE:1454123 3'  
DEFINITION similar to contains Alu repetitive element; contains element MER22  
repetitive element ;, mRNA sequence.

ACCESSION AA398331  
VERSION AA398331.1 GI:3047736  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 151)  
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:2043584.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
unknown library type  
Seq primer: -40ml3 fwd. ET from Amersham.  
Location/Qualifiers  
1..151  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1454123"  
/clone\_lib="NCI\_CGAP\_Ov8"  
/tissue\_type="serous adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: ovary; Vector: pCMV-Sport4; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.1 kb. Life Technologies catalog #:  
10982-015"

BASE COUNT 25 a 54 c 37 g 34 t 1 others

ORIGIN

Query Match 0.9%; Score 38; DB 40; Length 151;  
Best Local Similarity 100.0%; Pred. NO. 0.00028;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 1931  
|||||  
Db 63 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 100

RESULT 45  
AA397419 157 bp mRNA EST 13-AUG-1997  
LOCUS nc64f11.s1 NCI\_CGAP\_Prl Homo sapiens cDNA clone IMAGE:771405  
DEFINITION similar to contains Alu repetitive element.;, mRNA sequence.

ACCESSION AA397419  
VERSION AA397419.1 GI:2050432

KEYWORDS EST.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 157)  
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:638990.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Harston Linehan, M.D., Rodrigo Chuaqui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 656 Std Error: 0.00  
Seq primer: -41ml3 fwd. ET from Amersham  
High quality sequence stop: 155.  
Location/Qualifiers  
1..157  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:771405"  
/clone\_lib="NCI\_CGAP\_Prl"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected, histologically normal  
prostate epithelial cells. Double-stranded cDNA was  
ligated to EcoRI adaptors. 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the upc-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Krizman."  
28 a 50 c 39 g 40 t

BASE COUNT 28 a 50 c 39 g 40 t

ORIGIN

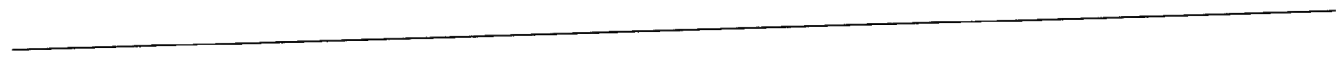
Query Match 0.9%; Score 38; DB 33; Length 157;  
Best Local Similarity 100.0%; Pred. NO. 0.00027;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 1931  
|||||  
Db 63 CAACCTCTGCTCCCGGGTTCAGTGATTCCTCGCT 100

Search completed: April 6, 2000, 00:07:41  
Job time: 19739 sec

us-09-090-672b-1.oligo\_1.rst

Fri Apr 7 09:12:41 2000





T18551;  
 06-MAY-1997 (first entry)  
 Human polycystic kidney disease normal PKD1 gene.  
 Adult onset polycystic kidney disease; APKD; autosomal dominant;  
 mutant; transversion; transition; deletion; insertion; ds.  
 Homo sapiens.  
 Key Location/Qualifiers  
 misc\_feature 4379..5272  
 /tag= a  
 /note= "Specifically claimed region of intronless  
 cDNA identified by exon trapping"  
 old\_sequence replace(50652..50653, cg)  
 /tag= b  
 /note= "changes Val codon to Leu codon"  
 old\_sequence replace(50796..50797, cg)  
 /tag= c  
 /note= "replaces Val codon by Leu codon"  
 old\_sequence insert(51827..51828, cc)  
 /tag= d  
 /note= "insertion, results in frameshift"  
 WO9612033-Al.  
 25-APR-1996.  
 11-OCT-1995; U13357.  
 12-OCT-1994; US-323443.  
 31-JAN-1995; US-381520.  
 (IGIG-) IG LAB INC.  
 (UYJO ) UNIV JOHNS HOPKINS.  
 Burn TC, Connors TD, Dackowski W, Germino G, Klingner KW;  
 Landes GM, Qian F;  
 WPI; 96-222017/22.  
 Isolated human polycystic kidney disease gene and its mutants  
 useful for treatment of polycystic kidney disease and screening for  
 carriers  
 Claim 1; Fig 1; 65pp; English.  
 The present sequence is that of the normal human PKD1 gene from  
 chromosome 16. Mutations in this gene (e.g. transitions,  
 transversions, deletions and/or insertions) are associated with  
 adult-onset polycystic kidney disease (APKD). The PKD1 locus is  
 GC-rich (62.4%). Comparison of this sequence with a previously  
 reported partial cDNA sequence revealed differences at three  
 locations (see features table). The most significant difference is  
 the presence of two additional cytosine residues on the plus-strand  
 at position 4566 of the previously reported sequence. The insertion  
 results in a frame-shift in the predicted protein coding sequence,  
 leading to replacement of 92 C-terminal amino acids with a novel  
 12 amino acid C-terminus. The PKD1 gene contains 23 Alu repeats.  
 There is a region consisting of 17 tandem copies of a perfect 27 bp  
 repeat and two large CT-rich regions.  
 Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T;  
 Sequence 53577 BP; 8495 A; 17681 C; 15785 G; 11616 T;  
 Query Match 1.8%; Score 53; DB 1; Length 53577;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-11;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 794108 standard; DNA; 53577 BP.  
 T94108;  
 01-JUN-1998 (first entry)  
 Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.  
 Human; polycystic kidney disease 1; PKD1; treatment;  
 autosomal dominant polycystic kidney disease; APKD; ss.  
 Homo sapiens.  
 WO9744457-Al.  
 27-NOV-1997.  
 22-MAY-1997; U08799.  
 03-JUN-1996; US-658136.  
 62 CTCACTGCAAGCTCCGCTCCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCC 114  
 |||||||  
 7885 CTCACTGCAAGCTCCGCTCCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCC 7937  
 RESULT 14  
 ID T94108 standard; DNA; 53577 BP.  
 AC T94108;  
 DT 01-JUN-1998 (first entry)  
 DE Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.  
 KW Human; polycystic kidney disease 1; PKD1; treatment;  
 KW autosomal dominant polycystic kidney disease; APKD; ss.  
 OS Homo sapiens.  
 PN WO9744457-Al.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08799.  
 PR 03-JUN-1996; US-658136.  
 Query Match 1.8%; Score 53; DB 1; Length 53577;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-11;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 794108 standard; DNA; 53577 BP.  
 T94108;  
 01-JUN-1998 (first entry)  
 DE Human PKD1 locus between chromosomal markers ATP1 (ATP6C) and D16S84.  
 KW Human; polycystic kidney disease 1; PKD1; treatment;  
 KW autosomal dominant polycystic kidney disease; APKD; ss.  
 OS Homo sapiens.  
 PN WO9744457-Al.  
 PD 27-NOV-1997.  
 PF 22-MAY-1997; U08799.  
 PR 03-JUN-1996; US-658136.

PR 24-MAY-1996; US-655360.  
 PA (GENZ ) GENZYME CORP.  
 PI Burn T, Connors T, Dackowski W, Germino G, Klingner K,  
 Qian F;  
 WPI; 98-018511/02.  
 Human polycystic kidney disease gene, PKD1 - useful to treat and  
 diagnose human autosomal or adult onset polycystic kidney disease  
 Example 5; Pages 60-89; 257pp; English.  
 The present sequence is the human polycystic kidney disease 1  
 (PKD1) locus between chromosomal markers ATP1 (ATP6C) and D16S84.  
 The PKD1 gene or polypeptide may be used to treat autosomal  
 dominant polycystic kidney disease (APKD), and identify carriers  
 of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies  
 (Ab) that distinguish between normal and mutant PKD1 sequences can  
 also be used in diagnostic tests. Anti-PKD1 Ab may also be used to  
 perform subcellular and histochemical localisation studies, and to  
 block the function of PKD1. Ab are also useful in rational drug  
 design studies to identify and test inhibitors of PKD1. Sense and  
 antisense sequences derived from the PKD1 gene may be used for  
 detection and therapy.  
 Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T;  
 Query Match 1.8%; Score 53; DB 1; Length 53577;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-11;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 62 CTCACTGCAAGCTCCGCTCCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCC 114  
 |||||||  
 7884 CTCACTGCAAGCTCCGCTCCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCC 7936  
 RESULT 15  
 ID Q12223/c  
 AC Q12223 standard; DNA; 1474 BP.  
 DT 13-SEP-1991 (first entry)  
 DE Erbb-3 gene clone E3-1 v-erbB homologous region.  
 KW Epidermal growth factor; EGFR; receptor; erbB; antibodies; tumour;  
 v-erbB; probe; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT exon 66..221  
 /tag= a  
 FT exon 657..855  
 /tag= b  
 FT exon 973..1446  
 /tag= c  
 WO9108214-A.  
 13-JUN-1991.  
 30-NOV-1990; U07025.  
 01-DEC-1989; US-444406.  
 (USDC ) US SEC OF COMMERCE.  
 (USSH ) NAT INST OF HEALTH.  
 Kraus MH, Aaronson SA;  
 WPI; 91-193144/26.  
 P-PSDB; R12607.  
 Epidermal growth factor receptor related gene - used for  
 producing receptors and antibodies for detecting and treating  
 tumours  
 Disclosure; Fig 3; 34pp; English.  
 The 1.5 Kbp region from the EcoRI to the PstI site shows the region  
 of homology in the human erbB-3 gene. The region contains 3 ORFs.  
 See Q12224 for full erbB-3 gene.  
 Sequence 1474 BP; 375 A; 353 C; 387 G; 359 T;  
 Query Match 1.7%; Score 52; DB 1; Length 1474;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 64 CACTGCAAGCTCCGCTCCCTGGGTTACGCCATTCTCCTGCCTCAGCCTCC 115  
 |||||||



Db 636 CACTGCAAGCTCCGCTCCTGGTTACAGCATTCTCTGCTCAGCCTCC 585

## RESULT 16

ID Q03369 standard; DNA; 1494 BP.  
 AC Q03369;  
 DT 20-DEC-1990 (first entry)  
 DE Sequence encoding human lysozyme  
 KW Transgenic animal; enzyme; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT signal\_peptide 14..67  
 FT mat\_peptide 68..460  
 FT /\*tag= a  
 FT /\*tag= b  
 PN J02005879-A.  
 PD 10-JAN-1990.  
 PF 21-JUN-1988; 151106.  
 PR (TAXE) Takeda Chemical Ind KK.  
 PA WPI 90-053916/08.  
 DR P-PSDB; R06108.  
 PT Human lysozyme gene -  
 PS useful promoter for genetic expression of animal culturing cell  
 PS ; Fig 4-1 to 4-3 Pages 834-835; 13pp; Japanese.  
 CC Human lysozyme DNA can be used to obtain transgenic animals, e.g. mice.  
 CC It can also be produced by culturing animal cells, and it can  
 CC promote the genetic expression of these animal cells.  
 SQ Sequence 1494 BP; 441 A; 298 C; 309 G; 446 T;

Query Match 1.7%; Score 52; DB 1; Length 1494;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 140 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGGTTTCACC 191  
 Db 890 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGGTTTCACC 941

## RESULT 17

ID N92386 standard; DNA; 1496 BP.  
 AC N92386;  
 DT 03-JUN-1990 (first entry)  
 DE DNA coding for the signal peptide and mature protein of human lysozyme  
 DE (HL)  
 KW Human lysozyme; HL; bacterial infection.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT signal\_peptide 14..67  
 FT mat\_peptide 68..457  
 FT /\*tag= a  
 FT /\*tag= b  
 FT /\*tag= c  
 PN J01074989-A.  
 PD 20-MAR-1989.  
 PF 16-SEP-1987; 229752.  
 PR (TAXE) Takeda Chemical Ind KK.  
 PA WPI 89-127529/17.  
 DR P-PSDB; P93510.  
 PT New DNA sequence -  
 PT encoding signal peptide and mature protein of human lysozyme  
 PS Figure 2-1 - 2-1; pages 12-13; 13pp; Japanese.  
 CC It is useful for effective expression of human lysozyme in yeast or  
 CC animal cells. Also, DNA encoding various proteins can be linked to the  
 CC 3' end of the HL signal peptide so that it is downstream of the  
 CC promoter of the expression vector. Expression of such proteins is  
 CC possible in E. coli, Bacillus subtilis, yeast or animal cells. Lysozyme

CC is useful for controlling bacterial infection and unlike chicken lysozyme  
 CC HL has little adverse effects.  
 SQ Sequence 1496 BP; 443 A; 299 C; 308 G; 446 T;

Query Match 1.7%; Score 52; DB 1; Length 1496;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-10;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 140 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGGTTTCACC 191  
 Db 890 GCCACACGCCCGCTAATTTTGTATTTTAGTAGACAGGTTTCACC 941

## RESULT 18

ID T10060/c standard; DNA; 1542 BP.  
 AC T10060;  
 DT 13-MAY-1996 (first entry)  
 DE erbB-3 genomic sequence.  
 KW tumour; antibody; proto-oncogene; cancer; receptor tyrosine kinase;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT exon 66..221  
 FT /\*tag= a  
 FT /codon\_start= 66..68  
 FT Intron 222..779  
 FT /\*tag= b  
 FT exon 778..855  
 FT /\*tag= c  
 FT Intron 856..1039  
 FT /\*tag= d  
 FT exon 1040..1185  
 FT /\*tag= e  
 PN US5480968-A.  
 PD 02-JAN-1996.  
 PF 04-DEC-1989; 444406.  
 PR 04-DEC-1989; US-444406.  
 PR 10-NOV-1992; US-978895.  
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.  
 PI Aaronson SA, Kraus ME;  
 DR WPI 96-068302/07.  
 DR P-PSDB; R88459.  
 PT erbB-3 receptor protein binding site peptide and specific antibody -  
 PT useful in assays to detect the expression of erbB proto-oncogene(s)  
 PS Example 2; Column 27-30; 41pp; English.  
 CC A human DNA fragment (T10060) codes for a portion (R88459) of a new  
 CC functional member of the erbB proto-oncogene family, designated  
 CC erbB-3. The DNA was identified by screening a human genomic DNA  
 CC with a probe specific for an oncogenic vital form of the erbB  
 CC gene, v-erbB, and was subsequently cloned from a human thymus  
 CC genomic library. The erbB-3 gene was mapped to 12q13. Elevated  
 CC levels of erbB-3 mRNA were demonstrated in certain mammary tumour cell  
 CC lines. The gene, or corresponding full-length cDNA (see T10059), can  
 CC be used to prepare erbB-3 or to detect expression of erbB-3 genes in  
 CC normal and tumour tissues.  
 SQ Sequence 1542 BP; 389 A; 371 C; 397 G; 385 T;

Query Match 1.7%; Score 52; DB 1; Length 1542;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 CACTGCAAGCTCCGCTCCTGGTTACAGCATTCTCTGCTCAGCCTCC 115  
 Db 636 CACTGCAAGCTCCGCTCCTGGTTACAGCATTCTCTGCTCAGCCTCC 585

## RESULT 19

ID V58735/c standard; cDNA; 1542 BP.  
 AC V58735;





Veterinary School, 3800 Spruce Street, Philadelphia, PA 19104-6008,  
USA

# FEATURES

source

Location/Qualifiers  
1. .2253  
/organism="Hyalophora cecropia"  
/db\_xref="taxon:7123"  
/tissue\_type="fat body"  
/dev\_stage="5th instar larva"  
10. .57  
/product="arylphorin"  
10. .2124  
/note="hemolymph aromatic amino acid rich storage protein;  
ArH"

sig\_peptide

CDS

/codon\_start=1  
/product="arylphorin precursor"  
/protein\_id="AAB86644.1"  
/db\_xref="GI:2625146"  
/translation="MTVLYLAGLIALVSSVSPVAHYKTKDVAFFVERQKKVLS  
LFQDVQVNDIEYIKIGDYDEANIDYNNKAAVEFLKLYRTGYLPKIEFSIFH  
HKLERAIALFHLFYAKDFEYKSAFAFVHNEGQFLYAYIAIQRNTHGFLV  
PAYEYVPOFVNMOTTVRIYTRKMDGILPKTAINIGIVKEEHYIYANYSNTEL  
YNNDEORLYTIDEDIGENSYYIYHSHLPFWTSGYGNLKHRRGEIYYFYQQLTR  
YPERLNTGLSIPESFNSPIKTYGYPLMTSYYPAPQENYLNHKNVYAEIRFL  
DIPEKTFVSLQGGFESNGKKIDPHDKALFNFGYQWQENADLYGEEYKDYQSYE  
IVARHVLGAAPKPDOKHIFMPSALDFYOTLRDPAFYQLYNRVGVINAFKHLVYP  
KSLVNAIEKSDVATDVLKIFMGPKIDNDGFPITLSENNKFEFLDNTFKHITPGQN  
KIVNNEERLIERKDSLELTLKLLDGRVYFDMSEEGIMPKRLMLPRGTGGTGP  
QFVYFIPSESSHDLTFEAFVIDNKLPLGYFPDPRVDSTSFKOPNMFPRDVFYIHG  
EHPFYKINYPSPISIKTNEVPKQ"

58. .2121  
/note="N-terminal sequence of the mature protein in the  
blood"  
/evidence="experimental"  
/product="arylphorin"  
717 a 482 c 414 g 640 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 32; DB 35; Length 2253;

Best Local Similarity 100.0%; Pred. No. 1.1e-05;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA... 2276

DB 2202 TTTATTTAAAAA... 2233

RESULT 23

AF032398

LOCUS

DEFINITION

AF032398 2431 bp mRNA INV 28-SEP-1999

Hyalophora cecropia moderately methionine rich hexamerin precursor,

mRNA, complete cds.

ACCESSION

AF032398

VERSION

AF032398.1

KEYWORDS

cecropia moth.

ORGANISM

Hyalophora cecropia

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;

Ditrysia; Bombycoidea; Saturniidae; Saturniinae; Hyalophora.

REFERENCE

1 (bases 1 to 2431)

2 (bases 1 to 2431)

3 (bases 1 to 2431)

4 (bases 1 to 2431)

5 (bases 1 to 2431)

6 (bases 1 to 2431)

7 (bases 1 to 2431)

8 (bases 1 to 2431)

9 (bases 1 to 2431)

10 (bases 1 to 2431)

TITLE

JOURNAL

USA

FEATURES

source

Location/Qualifiers  
1. .2431  
/organism="Hyalophora cecropia"  
/db\_xref="taxon:7123"  
/tissue\_type="fat body"  
/dev\_stage="5th instar larva"  
32. .82  
/product="moderately methionine rich hexamerin"  
32. .2290  
/note="moderately methionine rich fat body storage  
protein; MTH"

sig\_peptide

CDS

/codon\_start=1  
/product="moderately methionine rich hexamerin precursor"  
/protein\_id="AAB86646.1"  
/db\_xref="GI:2625150"  
/translation="MRAAVLLVGLAALASARPNDNDVNFVSMIDKQQLFLKLLNH  
VEVPYKDVSEIGNEKEDIMNELYTKTEVKSFMVLMKLVGLPRGEVFTLVNDQPM  
KEVVMHILYIAKDESTEIKTACWMLYLNSEFVLSVAVRVRECKVGLVPPPY  
ETYPYFVRAVDIQKALIKMRKGLDLKLSFEYGIKKTEDVYIIDENVDRVALN  
EDDKYETEDIDLTYYVYHVDVPMKDLFQKLTREELTYLYQQLLARYVL  
PELSGELIEPESFYPIQGYWPKLHNGIELPARNNVYSKYNTLISUYEE  
VETLKEALMGYIDMGIRLELTPEDETIGKMIYKVAKEQDSCKVAYRYLLI  
VMSAIGLNTLESNDKYFVAVPTVLDSTOTALRDPFYQLOKRLNIVQLFKRLPCYTN  
EELYPGKLENNVVVDKLLTYFDYFDMQNTNLTSEEEIKKASMSGFLVKRLRNH  
EPFYKLSLSDKAVDCVRIFFMGPKEDQGLRIDINKRLNFVEMDSFVYKLTGKN  
TIVRKSSMHLVPLKRDNTSDLWKVESITDMDLFDVQKLNRYFTGFTTRILLPGSV  
SLCPMVTVMVPLKLDVNDVMSALDPTSKALSEDFSTVLMKMPGLGFLDRIYDV  
TNFYTSNMKFVDVTIFHKNSCDMKMRKDWKLVKSNLASKPTIESNTYFVDTDLNSS  
SKVSSANVDM"

83. .2287

/note="alternate form of the mature protein in the blood"

/evidence="experimental"

/product="moderately methionine rich hexamerin"

86. .2287

/note="alternate form of the mature protein in the blood"

/evidence="experimental"

/product="moderately methionine rich hexamerin"

777 a 474 c 487 g 693 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 32; DB 35; Length 2431;

Best Local Similarity 100.0%; Pred. No. 1.1e-05;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA... 2276

DB 2399 TTTATTTAAAAA... 2430

mat\_peptide

mat\_peptide

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0;

QY 2245 TTTATTTAAAAA... 2276

DB 2399 TTTATTTAAAAA... 2430

RESULT 24

DDICNPA

LOCUS

DEFINITION

DDICNPA 6372 bp DNA INV 07-MAR-1995

Dictyosellum discoideum cyclic nucleotide phosphodiesterase gene,

complete cds.

ACCESSION

M23449

VERSION

M23449.1

KEYWORDS

cyclic nucleotide phosphodiesterase.

SOURCE

D. discoideum (strain NC4; cell line AX3-K) DNA, clones pGP-1 and

PC34

ORGANISM

Dictyosellum discoideum

Eukaryota; Dictyosteliida; Dictyostellium.

REFERENCE

1 (bases 1 to 6372)

2 (bases 1 to 6372)

3 (bases 1 to 6372)

4 (bases 1 to 6372)

5 (bases 1 to 6372)

6 (bases 1 to 6372)

7 (bases 1 to 6372)

8 (bases 1 to 6372)

9 (bases 1 to 6372)

10 (bases 1 to 6372)

11 (bases 1 to 6372)

12 (bases 1 to 6372)

13 (bases 1 to 6372)

14 (bases 1 to 6372)

15 (bases 1 to 6372)

16 (bases 1 to 6372)

17 (bases 1 to 6372)

18 (bases 1 to 6372)

19 (bases 1 to 6372)

20 (bases 1 to 6372)

21 (bases 1 to 6372)

22 (bases 1 to 6372)

23 (bases 1 to 6372)

24 (bases 1 to 6372)

25 (bases 1 to 6372)

26 (bases 1 to 6372)

27 (bases 1 to 6372)

28 (bases 1 to 6372)

29 (bases 1 to 6372)

30 (bases 1 to 6372)

31 (bases 1 to 6372)

32 (bases 1 to 6372)

33 (bases 1 to 6372)

34 (bases 1 to 6372)

35 (bases 1 to 6372)

36 (bases 1 to 6372)

37 (bases 1 to 6372)

38 (bases 1 to 6372)

39 (bases 1 to 6372)

40 (bases 1 to 6372)

41 (bases 1 to 6372)

42 (bases 1 to 6372)

43 (bases 1 to 6372)

44 (bases 1 to 6372)

45 (bases 1 to 6372)

46 (bases 1 to 6372)

47 (bases 1 to 6372)

48 (bases 1 to 6372)

49 (bases 1 to 6372)

50 (bases 1 to 6372)

51 (bases 1 to 6372)

52 (bases 1 to 6372)

53 (bases 1 to 6372)

54 (bases 1 to 6372)

55 (bases 1 to 6372)

56 (bases 1 to 6372)

57 (bases 1 to 6372)

58 (bases 1 to 6372)

59 (bases 1 to 6372)

60 (bases 1 to 6372)

61 (bases 1 to 6372)

62 (bases 1 to 6372)

63 (bases 1 to 6372)

64 (bases 1 to 6372)

65 (bases 1 to 6372)

66 (bases 1 to 6372)

67 (bases 1 to 6372)

68 (bases 1 to 6372)

69 (bases 1 to 6372)

70 (bases 1 to 6372)

71 (bases 1 to 6372)

72 (bases 1 to 6372)

73 (bases 1 to 6372)

74 (bases 1 to 6372)

75 (bases 1 to 6372)

76 (bases 1 to 6372)

77 (bases 1 to 6372)

78 (bases 1 to 6372)

79 (bases 1 to 6372)

80 (bases 1 to 6372)

81 (bases 1 to 6372)

82 (bases 1 to 6372)

83 (bases 1 to 6372)

84 (bases 1 to 6372)

85 (bases 1 to 6372)

86 (bases 1 to 6372)

87 (bases 1 to 6372)

88 (bases 1 to 6372)

89 (bases 1 to 6372)

90 (bases 1 to 6372)

91 (bases 1 to 6372)

92 (bases 1 to 6372)

93 (bases 1 to 6372)

94 (bases 1 to 6372)

95 (bases 1 to 6372)

96 (bases 1 to 6372)

97 (bases 1 to 6372)

98 (bases 1 to 6372)

99 (bases 1 to 6372)

100 (bases 1 to 6372)

101 (bases 1 to 6372)

102 (bases 1 to 6372)

103 (bases 1 to 6372)

104 (bases 1 to 6372)

105 (bases 1 to 6372)

106 (bases 1 to 6372)

107 (bases 1 to 6372)

108 (bases 1 to 6372)

109 (bases 1 to 6372)

110 (bases 1 to 6372)

111 (bases 1 to 6372)

112 (bases 1 to 6372)

113 (bases 1 to 6372)

114 (bases 1 to 6372)

115 (bases 1 to 6372)

116 (bases 1 to 6372)

117 (bases 1 to 6372)

118 (bases 1 to 6372)

119 (bases 1 to 6372)

120 (bases 1 to 6372)

121 (bases 1 to 6372)

122 (bases 1 to 6372)

123 (bases 1 to 6372)

124 (bases 1 to 6372)

125 (bases 1 to 6372)

126 (bases 1 to 6372)

127 (bases 1 to 6372)

128 (bases 1 to 6372)







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 09:00:38 ; Search time 345.99 seconds  
(without alignments)  
930.476 Million cell updates/sec

Title: US-09-090-672B-2  
Perfect score: 2689  
Sequence: 1 GTTGGAGTCTCTGGCGGCA.....AAAAAAAAAAAAAAAAAAAAA 2689

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 214294 seqs, 59861574 residues

Word size : 0

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/6.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PTUS9.COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	45	1.7	2989	7	5378464-1	Patent No. 5378464
2	45	1.7	3350	3	US-08-663-566A-1	Sequence 1, Appli
3	45	1.7	3350	4	US-08-023-610-1	Sequence 1, Appli
4	45	1.7	3350	4	US-08-288-065A-1	Sequence 1, Appli
5	45	1.7	3350	4	US-08-362-240A-1	Sequence 1, Appli
6	45	1.7	3350	6	PCT-US95-10245-1	Sequence 1, Appli
7	44	1.6	1391	1	US-08-261-662-1	Sequence 1, Appli
8	44	1.6	1391	6	PCT-US95-07752-1	Sequence 1, Appli
9	44	1.6	1781	1	US-08-314-615-2	Sequence 2, Appli
10	44	1.6	1781	1	US-08-314-362-2	Sequence 2, Appli
11	44	1.6	1781	1	US-08-433-010-2	Sequence 2, Appli
12	44	1.6	1781	2	US-08-245-295-4	Sequence 4, Appli
13	44	1.6	1781	2	US-08-481-130-4	Sequence 4, Appli
14	44	1.6	1781	2	US-08-656-984A-4	Sequence 2, Appli
15	44	1.6	1781	2	US-08-482-882-2	Sequence 2, Appli
16	44	1.6	1781	2	US-08-485-604-4	Sequence 2, Appli
17	44	1.6	1781	2	US-08-483-389-2	Sequence 2, Appli
18	44	1.6	1781	3	US-08-487-113D-2	Sequence 2, Appli
19	44	1.6	1781	3	US-08-487-595-4	Sequence 2, Appli
20	44	1.6	1781	3	US-08-473-503-2	Sequence 2, Appli
21	44	1.6	1781	3	US-08-483-932-2	Sequence 2, Appli
22	44	1.6	1781	4	US-08-720-420A-2	Sequence 1, Appli
23	44	1.6	2158	2	US-08-698-551-1	Sequence 1, Appli
24	44	1.6	2158	3	US-08-602-228-1	Sequence 1, Appli
25	44	1.6	2158	3	US-08-649-341A-1	Sequence 1, Appli
26	44	1.6	2158	3	US-08-494-440B-1	Sequence 1, Appli

27	44	1.6	2158	3	US-08-533-901B-1	Sequence 1, Appli
28	44	1.6	2158	3	US-08-339-032A-1	Sequence 1, Appli
29	44	1.6	2158	4	US-08-339-031A-1	Sequence 1, Appli
30	44	1.6	2158	6	PCT-US95-1274-1	Sequence 1, Appli
31	44	1.6	3471	6	PCT-US95-00227-2	Sequence 2, Appli
32	43	1.6	75	4	US-08-776-944-13	Sequence 13, Appli
33	43	1.6	80	2	US-07-920-281C-25	Sequence 25, Appli
34	43	1.6	635	2	US-08-455-633A-35	Sequence 35, Appli
35	43	1.6	635	2	US-08-416-336-5	Sequence 35, Appli
36	43	1.6	635	4	US-08-456-460C-35	Sequence 35, Appli
37	43	1.6	635	6	PCT-US94-05354-35	Sequence 35, Appli
38	43	1.6	688	7	5498694-3	Patent No. 5498694
39	43	1.6	1013	1	US-07-920-519-30	Sequence 30, Appli
40	43	1.6	1013	1	US-08-086-410-23	Sequence 23, Appli
41	43	1.6	1013	1	US-08-314-586-30	Sequence 30, Appli
42	43	1.6	1386	4	US-08-897-340-1	Sequence 1, Appli
43	43	1.6	1493	1	US-08-340-820-24	Sequence 24, Appli
44	43	1.6	1493	1	US-08-533-535-24	Sequence 24, Appli
45	43	1.6	1512	4	US-08-909-965C-8	Sequence 8, Appli

## ALIGNMENTS

RESULT 1  
5378464-1  
; Patent No. 5378464  
; APPLICANT: MCVEY, RODGER P.  
; TITLE OF INVENTION: MODULATION OF INFLAMMATORY RESPONSES  
; BY ADMINISTRATION OF GMP-140 OR ANTIBODY TO GMP-140  
; NUMBER OF SEQUENCES: 32  
; CURRENT APPLICATION NUMBER: US/07/320.408  
; FILING DATE: 08-MAR-1989  
; SEQ ID NO:1  
; LENGTH: 2989  
5378464-1

Query Match 1.7%; Score 45; DB 7; Length 2989;  
Best Local Similarity 100.0%; Pred. No. 2.6e-09;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689  
|||||  
DB 2911 ccc 2955

RESULT 2  
US-08-663-566A-1  
; Sequence 1, Application US/08663566A  
; Patent No. 5853733  
; GENERAL INFORMATION:  
; APPLICANT: Cochran, Mark D  
; APPLICANT: Macdonald, Richard D  
; TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys  
; TITLE OF INVENTION: and Uses Thereof  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: John P. White  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/663,566A  
; FILING DATE: June 13, 1996



94

### FEATURES

#### ADDRESS: 1. Location/Qualifiers

primer\_bind

BASE COUNT	105 a	77 c	104 g	80 t	4 others
ORIGIN					

Query Match 1.88; Score 27; DB 13; Length 370;

1416 TGGGCAACAGAGCAAGACCCAGTCTCA 1442

G07947/c				
LOCUS	G07947	471 bp	DNA	
DEFINITION	human STS CHLC.CTT16.P18204 clone CTT16, sequence tagged site.		STS	05-FEB-1997

VERSION  
KEYWORDS  
G07547:1 GI:536457  
STS: STS sequence: primer: sequence tagged site.

**SOURCE**  
human vector-pJcpl host=E.coli dut+ung+ (DH10B) Marker Selected  
E.coli sequence primer= sequence tagged site.  
genomic DNA prepared from XY individual of French nationality.

Eukaryota; Metazoa; Chordata; Chordata; Chordata

**TITLE**  
**JOURNAL**

UNPUBLISHED (1995)  
 Synonyms: CTT16, CHLC.CTT16.#T9987  
 Contact: Dr. Jeffrey C. Murray  
 UofI

The University of Iowa  
Department of Pediatrics, Iowa City, IA 52242, USA  
Tel: (319) 356-3508  
Fax: (319) 356-3347

Fax: (313) 350-3347  
Email: [teff-murray@uiowa.edu](mailto:teff-murray@uiowa.edu)

Primer A: TGGCTAGTGGACAAAGTGGT  
Primer B: CCCTTGTTATTACTGTTTCTT

SRS size: 174  
 PCR profile:
 

denature:	30 seconds at 94 degrees C
annealing:	75 seconds at 55 degrees C
extension:	15 seconds at 72 degrees C

Protocol:

extension: 6 minutes at 72 degrees C

PCR cycles: 27

extension: 10 seconds at 72 degrees C

Template: 30ng genomic DNA  
Primer: each 1.5 pmole  
dNTPs: each 200 uM

**Taq Polymerase:** 0.3 units

Total Vol:	10 ul
------------	-------

Buffer:

MgCl<sub>2</sub>: 1.5mM

KCl: 50mM  
Tris: 10mM

THIS: TOWNS

[illegible][illegible]

\_\_\_\_\_









**TITLE**  
**JOURNAL**  
**MEDLINE**  
**COMMENT**

Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dinke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Korak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,C.L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Hasetline,W.A., Fields,C., Fraser,C.M. and Venter,J.C.  
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)  
 12140200  
 On Dec 3, 1996 this sequence version replaced gi:1119186.  
 other ESTs: THC126123  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M13 Reverse.

**FEATURES**

source

Location/Qualifiers  
 1..274  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI-CGAP-GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI). Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATGAACTGGAGCGCCCTCATTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT**  
**ORIGIN**

87 a 51 c 53 g 81 t 2 others  
 Query Match 27.28; Score 74.6; DB 32; Length 274;  
 Best Local Similarity 98.7%; Pred. No. 7.4e-08;  
 Matches 74; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGCCATATATCAATACCAACTTGATAAATCTAGATTCCTGTCGTCTATAGAC 108  
 Db 1 TGCCATATATCAATACCAACTTGATAAATCTAGATTCCTGTCGTCTATAGAC 60  
 QY 109 CATGTTGTAGTAGG 123  
 Db 61 CATGTTGTAGTAGG 75

**RESULT 11**  
**LOCUS**  
**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**

AA488906 559 bp mRNA EST 15-AUG-1997  
 aa55a02.r1 NCI-CGAP-GCB1 Homo sapiens cDNA clone IMAGE:824810 5' similar to TR:G607003 G607003 BETA TRANSUCIN-LIKE PROTEIN. ;, mRNA sequence.  
 AA488906.1 GI:2218508  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 559)  
 Soares,M.B., Bonaldo,M.F., Jelenc,P., Su,L., Lawton,L and Efstathiadis,A.  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute. Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1397871.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Soares,M.B., Bonaldo,M.F., Jelenc,P., Su,L., Lawton,L and Efstathiadis,A.  
 Construction and characterization of a normalized cDNA library Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)  
 95023884  
 Contact: Bento Soares  
 Columbia University  
 Department of Psychiatry, 722 W 168th Street, Unit #41, New York, NY 10032

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: similarity on wrong strand  
 Seq primer: -28ml3 revl ET from Amersham  
 High quality sequence stop: 446.

Location/Qualifiers  
 1..559  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:824810"  
 /clone\_lib="NCI-CGAP-GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI). Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATGAACTGGAGCGCCCTCATTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT**  
**ORIGIN**

158 a 139 c 100 g 162 t  
 Query Match 23.68; Score 64.6; DB 34; Length 559;  
 Best Local Similarity 98.5%; Pred. No. 1.3e-05;  
 Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 59 ACAATACCAACTTCATTAATCTAGATTCCTGTCGTCTATAGACCACTGTTGTA 118  
 Db 559 ACAATACCAACTTCATTAATCTAGATTCCTGTCGTCTATAGACCACTGTTGTA 500  
 QY 119 GTAGG 123  
 Db 499 GTAGG 495

**RESULT 12**  
**LOCUS**  
**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**

T10350 259 bp mRNA EST 13-APR-1994  
 b4HB3MA-COT8-HAP-Ft Homo sapiens cDNA clone  
 b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.  
 T10350  
 T10350.1 GI:471699  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 259)  
 Soares,M.B., Bonaldo,M.F., Jelenc,P., Su,L., Lawton,L and Efstathiadis,A.  
 Construction and characterization of a normalized cDNA library Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)  
 95023884  
 Contact: Bento Soares  
 Columbia University  
 Department of Psychiatry, 722 W 168th Street, Unit #41, New York, NY 10032





4

6

8



BASE COUNT     174 A    123 C    96 G    203 T

ORIGIN

Query Match               23.2% ; Score 63.6; DB 104; Length 596;  
Best Local Similarity    94.3%; Pred. No. 2.e-05;  
Matches    66; Conservative    0; Mismatches    4; Indels      0; Gaps      0;

QY    1   CGTTACAGATCTTCGCGCTGGGTGAACACTACAAGGATCGTGCCTATATCAC 60  
||||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db    527 CGTTACAGATCTTCGCGCAGCGGTGAACACTACAAGGATCGTGCCTATATCAC 586  
  
QY    61 AAATACCAAAC 70  
|||||||  
Db    587 AAATACCAAAC 596

RESULT 14

LOCUS          Al196903

DEFINITION    ui55c03.yl Sugano mouse liver mlia Mus musculus cDNA clone  
IMAGE:1886308 5' similar to TR:O14432 O14432 TUPL.; mRNA  
sequence.

ACCESSION      Al196903

VERSION        Al196903

KEYWORDS       EST.

SOURCE        Al196903.1 GI:3749509

ORGANISM       house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 449)  
Marras,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T.,  
Geisel,S., Kucaba,T., Lacy,M.M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Watson,R.R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel.: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGJ:970632

Seq primer: custom primer used  
High quality sequence stop: 448.

FEATURES  
Source  
1..449  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1886308"  
/clone\_lib="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: pME18S-FL3; Site\_1: DraIII  
(CACTGTGTG); Site\_2: DraII (CAGCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGCCCTTTTTTTTTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTTGCCCTACTGG], digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACTGTGTG, 3' site CAGCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5Kb library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTCTGTCTATAAAGCTGG and 3' end  
primer CGACTCCAGCTCGAGACA."





PA (GENZ ) GENZYME CORP.  
 PI Burn T, Connors T, Dackowski W, Germino G, Klinger K,  
 PI Qian F;  
 DR WPI: 98-018511/02.  
 PT Human polycystic kidney disease gene, PKD1 - useful to treat and  
 PT diagnose human autosomal or adult onset polycystic kidney disease  
 PS Example 5; Pages 60-89; 257pp; English.  
 CC The present sequence is the human polycystic kidney disease 1  
 CC (PKD1) locus between chromosomal markers ATP6C and D16S84.  
 CC The PKD1 gene or polypeptide may be used to treat autosomal  
 CC dominant polycystic kidney disease (APKD), and identify carriers  
 CC of mutant PKD1 genes, i.e. subjects susceptible to APKD. Antibodies  
 CC (Ab) that distinguish between normal and mutant PKD1 sequences can  
 CC also be used in diagnostic tests. Anti-PKD1 Ab may also be used to  
 CC perform subcellular and histochemical localisation studies, and to  
 CC block the function of PKD1. Ab are also useful in rational drug  
 CC design studies to identify and test inhibitors of PKD1. Sense and  
 CC antisense sequences derived from the PKD1 gene may be used for  
 CC detection and therapy. 8495 A; 17684 C; 15782 G; 11616 T;  
 SQ Sequence 53577 BP; 8495 A; 17684 C; 15782 G; 11616 T;  
 Query Match 1.4%; Score 45; DB 1; Length 53577;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-10;  
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3110 GAGACGAGCTGACCAACATGCTGAAACCCGCTCTACTATAAAA 3154  
 DB 1692 GAGACGAGCTGACCAACATGCTGAAACCCGCTCTACTATAAAA 1648  
 RESULT 13  
 ID X02711 standard; DNA: 1083 BP.  
 AC X02711.  
 DT 14-MAY-1999 (first entry)  
 DE Human transaldolase TAL-H TARE-6 retrotransposable element.  
 DE Transaldolase; TAL-H; autoantigen; human; antibody; neuroprotector;  
 KW neurodegenerative autoimmune disease; multiple sclerosis; detection;  
 KW proliferation inhibitor; cytotoxicity inhibitor; binding inhibitor;  
 KW retrotransposon; retrotransposable element; TARE-6; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT misc\_signal 638..929  
 FT /\*tag= a  
 FT /note= "Alu-sc-like dimer"  
 FT US5879909-A.  
 PN 09-MAR-1999.  
 PD 09-OCT-1998; 057762.  
 PR 19-OCT-1994; US-326119.  
 PR 09-APR-1998; US-057762.  
 PA (UYNV ) UNIV NEW YORK STATE RES FOUND.  
 PI Perl A;  
 DR WPI: 99-203948/17.  
 PT Isolated human transaldolase gene - useful for raising antibodies  
 PT for detecting neurodegenerative autoimmune diseases, especially  
 PT multiple sclerosis  
 PS Claim 12; Fig 19; 59pp; English.  
 CC This sequence represents a human transaldolase (TAL-H) retrotransposable  
 CC element TARE 6. The TAL-H protein can be used in methods to raise  
 CC antibodies for detecting human transaldolase-mediated neurodegenerative  
 CC autoimmune diseases, especially multiple sclerosis. The protein has  
 CC neuroprotective properties and is a proliferation, cytotoxicity and  
 CC binding inhibitor.  
 SQ Sequence 1083 BP; 237 A; 302 C; 252 G; 292 T;  
 Query Match 1.3%; Score 44; DB 1; Length 1083;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2566 GGGGTTTCACTGTTAGCCAGGATGCTTGATCTCTGACCT 2609  
 |||||||||||||||||||||||||||||||||||||||||||||||||

DB 822 GGGTTTCACTGTTAGCCAGGATGCTTGATCTCTGACCT 865  
 RESULT 14  
 V57903/c  
 ID V57903 standard; DNA: 237326 BP.  
 AC V57903.  
 DT 21-DEC-1998 (first entry)  
 DE Hereditary haemochromatosis subregion from an HH affected individual.  
 DE Bovine butyrophilin; BF; human hereditary haemochromatosis; HFE;  
 KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;  
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;  
 KW type 1 sodium transport gene; ss.  
 OS Homo sapiens.  
 PN WO9814466-A1.  
 PD 09-APR-1998.  
 PF 30-SEP-1997; D17658.  
 PR 07-MAY-1997; US-852495.  
 PR 01-OCT-1996; US-724394.  
 PA (PROG-) PROGENITOR INC.  
 PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,  
 PI Tsuchihashi Z, Wolff RK;  
 DR WPI: 98-240014/21.  
 PT Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism  
 PS Claim 1; Fig 9; 209pp; English.  
 CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an hereditary  
 CC haemochromatosis (HH) affected individual. Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a RoRet gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.  
 SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;  
 Query Match 1.3%; Score 44; DB 1; Length 237326;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-09;  
 Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 3111 AGACGAGCTGACCAACATGCTGAAACCCGCTCTCTACTATAAAA 3154  
 |||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 49851 AGACGAGCTGACCAACATGCTGAAACCCGCTCTCTACTATAAAA 49808  
 RESULT 15  
 Q62613/c  
 ID Q62613 standard; DNA: 1688 BP.  
 AC Q62613.  
 DT 17-JAN-1995 (first entry)  
 DE Human mdr-1 promoter fragment.  
 DE Inducible promoter; cytostatic agent; adriamycin; vincristine;  
 KW multiple drug resistance; mdr-1; mammalian expression vector;  
 KW cancer therapy; ss.  
 OS Homo sapiens.  
 PN DE423878-A.  
 PD 19-MAY-1994.  
 PF 12-NOV-1992; 238778.  
 PR 12-NOV-1992; DE-238778.  
 PA (DELB-) DELBROECK CENT MOLEKULARE MEDIZIN MAX.



PI Stein U, Walther W;  
DR WPI: 94-168680/21.  
PT New mammalian expression vector useful for gene therapy, -  
PT comprising the promoter and/or the enhancer of the mdr-1 gene  
PT which is susceptible to cytostatic agents.  
PS Claim 2; Page 5; 7pp; German.  
CC This is a preferred fragment of the mdr-1 gene promoter, isolated  
CC from human DNA. The mdr-1 promoter and enhancer elements are  
CC inducible by cytostatic agents such as vincristine and adriamycin  
CC which are used in cancer therapy. Vectors comprising the promoter  
CC and enhancer sequences operably linked to heterologous genes coding  
CC for therapeutic agents are claimed. Using the vectors, the  
CC therapeutic agents (e.g. anti-oncogenes, interferon, IL-2, IL-6,  
CC etc.) can be expressed in cancer cells under the control of  
CC cytostatic agents.  
SQ Sequence 1688 BP; 556 A; 311 C; 371 G; 450 T;

Query Match 1.3%; Score 43; DB 1; Length 1688;  
Best Local Similarity 100.0%; Pred. No. 3.5e-09;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2567 GGGTTTCACGTGTAGCCAGGATGCTGTGATCTCTCCTGACCT 2609

DB 272 GGGTTCACGTGTAGCCAGGATGCTGTGATCTCTCCTGACCT 230

## RESULT 16

ID Q62612 standard; DNA; 2090 BP.  
AC Q62612;  
DT 17-JAN-1995 (first entry)  
DE Human mdr-1 promoter fragment.  
KW Inducible promoter; cytostatic agent; adriamycin; vincristine;  
KW multiple drug resistance; mdr-1; mammalian expression vector;  
KW cancer therapy; ss.  
OS Homo sapiens.  
PN DE4238778-A.  
PD 19-MAY-1994.  
PF 12-NOV-1992; 238778.  
PR 12-NOV-1992; DE-238778.  
PA (DELB-) DELBRUCK CENT MOLEKULARE MEDIZIN MAX.  
PI Stein U, Walther W;  
DR WPI: 94-168680/21.  
PT New mammalian expression vector useful for gene therapy, -  
PT comprising the promoter and/or the enhancer of the mdr-1 gene  
PT which is susceptible to cytostatic agents.

PS Claim 2; Page 4; 7pp; German.  
CC This is a preferred fragment of the mdr-1 gene promoter, isolated  
CC from human DNA. The mdr-1 promoter and enhancer elements are  
CC inducible by cytostatic agents such as vincristine and adriamycin  
CC which are used in cancer therapy. Vectors comprising the promoter  
CC and enhancer sequences operably linked to heterologous genes coding  
CC for therapeutic agents are claimed. Using the vectors, the  
CC therapeutic agents (e.g. anti-oncogenes, interferon, IL-2, IL-6,  
CC etc.) can be expressed in cancer cells under the control of  
CC cytostatic agents.  
SQ Sequence 2090 BP; 714 A; 365 C; 444 G; 567 T;

Query Match 1.3%; Score 43; DB 1; Length 2090;  
Best Local Similarity 100.0%; Pred. No. 3.5e-09;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2567 GGGTTTCACGTGTAGCCAGGATGCTGTGATCTCTCCTGACCT 2609

DB 674 GGGTTCACGTGTAGCCAGGATGCTGTGATCTCTCCTGACCT 632

## RESULT 17

OY Q81139  
ID Q81139 standard; DNA; 15328 BP.  
DT 17-JAN-1995 (first entry)

DE HPLA2-8 gene.  
KW HPLA2-8; phospholipase A2; PLA2; Batten disease;  
KW neuronal ceroid lipofuscinosis; gene therapy; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT exon 399..466  
FT /\*tag= a  
FT /\*label= Exon-1  
FT 2633..2777  
FT /\*tag= b  
FT /\*label= Exon-2  
FT 13862..14028  
FT /\*tag= c  
FT /\*label= Exon-3  
PN W09502328-A.  
PD 26-JAN-1995.  
PF 15-JUL-1994; U07926.  
PR 15-JUL-1994; US-091941.  
PR 26-JUL-1993; US-097354.  
PA (INCY-) INCYTE PHARM INC.  
PA (INDV) UNIV INDIANA FOUND.  
PI Seilhamer JJ, Tischfield JA;  
DR WPI: 95-067096/09.  
DR P-PSDB: R63052; R63053.  
DR Novel type III and IV low mol. wt. phospholipase A2 enzymes -  
PT from humans and rats, also nucleic acid sequences useful, e.g.  
PT for recombinant prodn. of enzymes, research into Batten's  
PT disease, etc.  
PS Claim 8; Page 67-76; 160pp; English.  
CC A human PLA2-encoding CDNA (Q81138) expressing HPLA2-10, was  
CC isolated from human brain RNA by RACE-PCR. 2 rat PLA2 cDNAs,  
CC designated RPLA2-8 (Q81136) and RPLA2-10 (Q81137), were isolated from  
CC rat brain and heart CDNA libraries, respectively. A partial human  
CC genomic counterpart to RPLA2-8, HPLA2-8 (Q81139), was also obtained.  
CC RPLA2-8 and HPLA2-8 have been designated type III PLA2, and RPLA2-10  
CC and HPLA2-10 as type IV. 3885 A; 3789 C; 4082 G; 3572 T;  
SQ Sequence 15328 BP; 3885 A; 3789 C; 4082 G; 3572 T;

Query Match 1.3%; Score 43; DB 1; Length 15328;  
Best Local Similarity 100.0%; Pred. No. 3.2e-09;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3231 TGAACCCGGAGCGAGGTTCGAGTGAGCGGAGATCACACCA 3273  
|||||  
DB 1790 TGAACCCGGAGCGAGGTTCGAGTGAGCGGAGATCACACCA 1832  
|||||

RESULT 18  
ID Q61108/c  
AC Q61108 standard; DNA; 223 BP.  
DT 16-MAR-1994 (first entry)  
DE Human brain Expressed Sequence Tag EST02702.  
DE Gene transcription product; genetic markers; tagging; in vivo;  
KW transcription; mapping; locations; chromosomes; Chromosomal; ss.  
OS Homo sapiens.  
PN W09316178-A.  
PD 19-AUG-1993.  
PF 12-FEB-1993; U01294.  
PR 12-FEB-1992; US-837195.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
PI Adams MD, Moreno RF, Venter CJ;  
DR WPI: 93-272882/34.  
DR Enriched oligonucleotides and corresp. sequences - used as  
PT markers for human genes transcribed in-vivo, facilitate tagging  
PT of most human genes  
PS Example 4; Page 440; 500pp; English.  
CC The Expressed Sequence Tag was isolated from a human brain CDNA  
CC library as part of a large set of ESTs which can be used as markers  
CC for human genes transcribed in vivo. They can be used to facilitate  
CC tagging of most human genes, for mapping locations of expressed genes



27782. 27861.29376. 29347.29456. 29570.29695. 29742.  
30178. 30207.30365. 30473.33547. 33605.34836. 34955.  
35525. 35643.))  
/gene="SAPK4"  
/note="match: proteins 014739 015124"  
/codon\_start=1  
/evidence="not\_experimental"  
/product="dj179N16.1 (Stress Activated Protein Kinase 4  
(Mitogen Activated Protein Kinase p38delta))"  
/protein\_id="CAB08438.1"  
/db\_xref="GI:3038774"  
/db\_xref="GI:3038774"  
/translation="MSLRKRGKGYKDVNTAWELPKTYVSPHVGSGAYGVCISAID  
KRSGEVAKKLSRPFQSEIFAKRAYRELLKMHQENIVIGLLDVTTPASSLRNFYD  
FYLVPFMDLQIMQMEFSEVKIOLVYOMLKLKLYHSAGVHRLDKPGLAVNE  
DCELKILDFGLARHADAEVTGYVYRWYRAPEVILSWMHYNQTVDIWSVGCIMAEMLT

Query Match 13.9% Score 19; DB 10; Length 172048;  
Best Local Similarity 100.0%; Pred. No. 1.5; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCACCGCAGCTGGCTGATG 19  
DB 41760 CCACCGCAGCTGGCTGATG 41742

RESULT 18  
LOCUS AC007158 204027 bp DNA HTG 26-OCT-1999  
DEFINITION Homo sapiens clone hRPK.90\_A\_1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2  
ordered pieces.  
ACCESSION AC007158  
VERSION AC007158.5 GI:5757580  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 204027)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Collangelo, M., Collins, S., Collymore, A.,  
Cooke, P., DeArrellano, K., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Kartas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,  
Marquis, N., McEwen, P., McGurk, A., McKernan, K., McLaughlin, J.,  
Melidrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 22, 1999 this sequence version replaced gi:4726136.  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.

\* the accession number will be preserved  
\* 1 160375: contig of 160375 bp in length  
\* 160376 204027: contig of 43652 bp in length.  
FEATURES  
source  
Location/Qualifiers  
1. 204027  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="hrpk.90\_A\_1"  
/clone\_lib="RPCI-11 Human Male BAC"  
BASE COUNT 63998 a 40107 c 39047 g 60875 t  
ORIGIN

Query Match 13.9% Score 19; DB 42; Length 204027;  
Best Local Similarity 100.0%; Pred. No. 1.5; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCACCGCAGCTGGCTGATG 19  
DB 111376 CCACCGCAGCTGGCTGATG 111394

RESULT 19  
LOCUS HUMVIPHMI 270 bp DNA PRI 14-JAN-1995  
DEFINITION Human vasoactive intestinal peptide and histidine-methionine amide  
peptide hormone (VIP/PHM-27) gene, exon 1.  
ACCESSION M14618  
VERSION M14618.1 GI:340266  
KEYWORDS histidine-methionine amide; hormone; peptide hormone; vasoactive  
intestinal peptide.  
SEGMENT 1 of 6  
SOURCE Human DNA (libraries of Lawn and EMBL), clones lambda-VIP-D and  
lambda-VIP-4.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 270)  
AUTHORS Linder, S., Barkhem, T., Norberg, A., Persson, H., Schalling, M.,  
Hokfelt, T. and Magnusson, G.  
TITLE Structure and expression of the gene encoding the vasoactive  
intestinal peptide precursor  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (2), 605-609 (1987)  
MEDLINE 87092456  
COMMENT Draft entry and clean copy of sequence [1] kindly provided by  
S.Linder, 02-MAR-1987.  
FEATURES  
Location/Qualifiers  
1. 270  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6q24-q27"  
prim\_transcript 100. 270  
/gene="VIP"  
intron 264. 270  
/note="VIP/PHM27 mRNA: G00-120-490"  
/gene="VIP"  
BASE COUNT 76 a 67 c 69 g 58 t  
ORIGIN Unreported.

Query Match 13.1% Score 18; DB 9; Length 270;  
Best Local Similarity 100.0%; Pred. No. 6.7; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 30 TGACTTCTTTTCAGAGGAC 47  
DB 22 TGACTTCTTTTCAGAGGAC 39



PS Example: Fig 2A-H; 98pp; English.  
 CC A cDNA molecule (T63334) codes for human nucleoprotein interactor  
 CC 1 (NPI-1) (W16327), a host cell protein which interacts with  
 CC influenza virus nucleoprotein (NP) and which may be an accessory  
 CC protein required for influenza virus replication. It is the human  
 CC homologue of the yeast protein SRP1. NPI-1 was identified by  
 CC interactive trap selection using LexA-NP as bait and yeast  
 CC transformed with a Hela cDNA library, and by 5' RACE (see also  
 CC T63341). NPI-1 to 6 cDNAs (see also T63335-39) have been  
 CC cloned and analysed. The human NSI1-1 gene (T63340), which codes  
 CC for a protein (W16329) that interacts with influenza virus protein  
 CC NS1 was similarly identified. The cDNA clones can be used to  
 CC produce human host cell proteins for use in screening for cpds.  
 CC that inhibit viral replication.  
 SQ Sequence 2940 BP; 826 A; 641 C; 627 G; 846 T;

Query Match 11.9%; Score 16; DB 1; Length 2940;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 ACAGAGATTAGACAAA 108  
 |||||  
 DB 766 ACAGAGATTAGACAAA 751

RESULT 6  
 X36424  
 ID X36424 standard; DNA; 536 BP.  
 AC X36424;  
 DT 06-JUL-1999 (first entry)  
 DE Human heregulin-like factor coding sequence clone HAGFE38R.  
 KW Human heregulin-like factor; HLF; cell growth regulator; diagnosis;  
 KW human system disorder; cancer; ss.  
 OS Homo sapiens.  
 PN W09857989-A1.  
 PD 23-DEC-1998.  
 PF 16-JUN-1998; U12403.  
 PR 17-JUN-1997; US-049942.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (GEU-) UNIV GEORGETOWN.  
 PI H1jazi MM, King CR, Ruben SM, Young P;  
 DR WPI; 99-095327/08.  
 PT New isolated heregulin-like factor - used to develop products for  
 PT the diagnosis and treatment of disorders involving regulation of  
 PT cell growth, particularly cancers  
 PS Claim 20; Page 90-91; 118pp; English.  
 CC This sequence encodes the human heregulin-like factor (HLF) of the  
 CC invention. The HLF is involved in the regulation of cell growth.  
 CC Detection of different levels of expression of the HLF gene can be used  
 CC for the diagnosis of disorders, e.g. in the neural system. In  
 CC particular, detection of different levels of HLF gene expression in cells  
 CC or body fluid of an individual can be used for diagnosing cancer. The  
 CC products can also be used in the treatment of disorders involving  
 CC abnormal levels of HLF activity.  
 SQ Sequence 536 BP; 141 A; 131 C; 125 G; 120 T;

Query Match 11.1%; Score 15; DB 1; Length 536;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 AGGGCTTCCCTTT 39  
 |||||  
 DB 472 AGGGCTTCCCTTT 486

RESULT 7  
 N80469  
 ID N80469 standard; cDNA; 713 BP.  
 AC N80469;  
 DT 06-DEC-1990 (first entry)  
 DE Gastrin inhibitory polypeptide precursor.

KW Gastrin inhibitory polypeptide precursor; GIP; probe; insulin;  
 OS diabetes; ss.  
 FH Key  
 FT cds  
 FT 101..562  
 FT Location/Qualifiers  
 FT /tag= a  
 FT /label=CIP precursor  
 FT /note="Claim 1"  
 FT 1..160  
 FT signal\_peptide  
 FT /tag= b  
 FT 254..379  
 FT mat\_peptide  
 FT /tag= c  
 FT /product=processed GIP  
 FT 290..303  
 FT other\_feature  
 FT /tag= d  
 FT /label=probe  
 FT /note="used to design probe for diabetes diagnosis"  
 FT 356..372  
 FT other\_feature  
 FT /tag= e  
 FT /label=probe  
 FT /note="used to design probe for diabetes diagnosis"  
 FT 553..658  
 FT polya\_signal  
 FT /tag= f  
 FT 692..697  
 FT polya\_signal  
 FT /tag= g  
 FT /note="most likely of the two signals"  
 FT BP-269072-A.  
 PD 01-JUN-1988.  
 PF 24-NOV-1987; 117325.  
 PR 27-NOV-1986; JP-282812.  
 PA (SANW) SANWA KAGAKU KENKYUSHO.  
 PI Takeda J, Imura H, Seino Y, Tanaka K, Takahashi H, Mitani T;  
 PI Kuroono M, Sawai K;  
 DR WPI; 88-148897/22.  
 DR P-PSDB; P80287.  
 PT DNA encoding human gastric inhibitory polypeptide precursor -  
 PT used as probe for diagnosis of diabetes and for producing  
 PT polypeptide(s) for diabetes treatment.  
 PS Disclosure; p; English.  
 CC The sequence was obt'd. from a clone isolated from a cDNA library  
 CC prepared from total RNA extracted from the duodenum of a patient  
 CC during panceato-duodenectomy for pancreatic cancer. The sequence  
 CC can be used to design probes (see features) for the diagnosis of  
 CC diabetes. It may also be ligated into an expression vector for  
 CC prodn. of recombinant GIP. GIP accelerates the gastric secretion  
 CC and insulin secretion and can be used to treat diabetes.  
 CC See also N82432 and N82433.  
 SQ Sequence 713 BP; 193 A; 187 C; 191 G; 142 T;

Query Match 11.1%; Score 15; DB 1; Length 713;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 AGAAGAAAGAGGTC 30  
 |||||  
 DB 165 AGAAGAAAGAGGTC 179

RESULT 8  
 V15706  
 ID V15706 standard; cDNA; 863 BP.  
 AC V15706;  
 DT 03-JUL-1998 (first entry)  
 DE Recombinant mosquito salivary allergen rae2 a 3 cDNA.  
 KW Recombinant mosquito; salivary allergen; rae2 a 3; determination;  
 KW bite sensitivity; epi-cutaneous test; skin test; intradermal test;  
 KW allergy diagnosis; immunotherapy; desensitisation; ss.  
 OS Aedes aegypti.  
 FT key  
 FT CDS  
 FT Location/Qualifiers  
 FT /tag= a  
 FT 12..773  
 FT sig\_peptide  
 FT 12..65





CC This is the MAGE-B cluster DNA sequence. The invention provides a novel  
 CC method for screening for the possibility of a testicular seminoma,  
 CC non-small cell lung carcinoma, melanoma, breast cancer, sarcoma, or  
 CC leukemia in a sample. The method comprises contacting the sample with  
 CC at least one nucleic acid molecule which hybridises to mRNA corresponding  
 CC to an MAGE-xp gene, and determining hybridisation as a determination of  
 CC possible presence of testicular seminoma, non-small cell lung carcinoma  
 CC melanoma, breast cancer, sarcoma, or leukaemia in the sample. By assaying  
 CC for the MAGE-B1 gene, or MAGE-B2 gene, both contained within this 40352  
 CC base pair MAGE-B cluster DNA sequence, the presence of such cancers in a  
 CC sample can be determined. The genomic DNA that encodes the MAGE-B2 gene  
 CC consists of nucleotides 3266-7979 of this MAGE-B cluster DNA sequence.  
 CC The nucleic acid molecule that encodes a MAGE-B1 gene, consists of, in  
 CC 5' to 3' order, nucleotides 31403-31474, 33958-34062, 35057-35139, and  
 CC 38088-39691, nucleotides 31403-31474, 33958-34062, and 38088-39691,  
 CC nucleotides 35057-35139 and 38088-39691, and nucleotides 33958-34062 and  
 CC 38088-39691 of this sequence. The primers specific for the MAGE-B  
 CC gene can be used in a kit to amplify a MAGE-B gene.  
 SQ Sequence 40352 BP; 11599 A; 8654 C; 9772 G; 10327 T;

Query Match 0.9%; Score 38; DB 1; Length 40352;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;  
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1909 GGGTCAAGTGATTCCTCGCTTAGCCTCCGAGTAG 1946

|||||

Db 8269 GGGTCAAGTGATTCCTCGCTTAGCCTCCGAGTAG 8232

#### RESULT 4

V57926  
 ID V57926 standard; DNA; 235033 BP.

AC 23-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an unaffected individual.  
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;  
 KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;  
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;  
 KW type 1 sodium transport gene; ss.

OS Homo sapiens.

PN WO9814466-A1.

PD 09-APR-1998.

PF 30-SEP-1997; U17658.

PR 07-MAY-1997; US-852495.

PR 01-OCT-1996; US-724394.

PA (PROG-) PROGENITOR INC.

PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,

PI Tsuchihashi Z, Wolff RK;

DR WPI; 98-240014/21.

PT Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism

PS Example 2; Fig 8; 209pp; English.

CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an individual  
 CC unaffected by hereditary haemochromatosis (HH). Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a RoRet gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.

SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

Query Match 0.9%; Score 38; DB 1; Length 235033;

Best Local Similarity 100.0%; Pred. No. 3.8e-05;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTCAAGTGATTCCTCGCT 1931

|||||

Db 208777 CAACCTCTGCTCCCGGGTCAAGTGATTCCTCGCT 208814

#### RESULT 5

V57903  
 ID V57903 standard; DNA; 237326 BP.

AC 21-DEC-1998 (first entry)

DE Hereditary haemochromatosis subregion from an HH affected individual.  
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;  
 KW diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;  
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;  
 KW type 1 sodium transport gene; ss.

OS Homo sapiens.

PN WO9814466-A1.

PD 09-APR-1998.

PF 30-SEP-1997; U17658.

PR 07-MAY-1997; US-852495.

PR 01-OCT-1996; US-724394.

PA (PROG-) PROGENITOR INC.

PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,

PI Tsuchihashi Z, Wolff RK;

DR WPI; 98-240014/21.

PT Hereditary haemochromatosis gene products - used to develop products  
 PT for the diagnosis and treatment of hereditary disorders in iron  
 PT metabolism

PS Claim 1; Fig 9; 209pp; English.

CC The present invention describes hereditary haemochromatosis gene  
 CC products from the human haemochromatosis gene. The present sequence  
 CC represents a hereditary haemochromatosis subregion from an hereditary  
 CC haemochromatosis (HH) affected individual. Also described is a  
 CC method to determine the presence or absence of the common hereditary  
 CC haemochromatosis (HFE) gene mutation in an individual comprising:  
 CC (a) providing DNA or RNA from the individual; and (b) assessing the  
 CC DNA or RNA for the presence or absence of a haplotype or genotype where  
 CC the presence or absence of the haplotype genotype indicates the likely  
 CC presence of the HFE gene mutation in the genome of the individual. The  
 CC HFE gene sequences from the present invention can be used to develop  
 CC products for use in the diagnosis and treatment of HFE. The present  
 CC invention also describes BTF genes, which are homologues of the milk  
 CC protein butyrophilin (BT), and can be used in the production of agonists  
 CC and antagonists of BT function. Also described are: (1) a RoRet gene  
 CC which can be used to develop products for the study, diagnosis and  
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes  
 CC which are homologues of a type 1 sodium transport gene, and can  
 CC similarly be used for hypophosphatemia.

SQ Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T;

Query Match 0.9%; Score 38; DB 1; Length 237326;

Best Local Similarity 100.0%; Pred. No. 3.8e-05;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1894 CAACCTCTGCTCCCGGGTCAAGTGATTCCTCGCT 1931

|||||

Db 83860 CAACCTCTGCTCCCGGGTCAAGTGATTCCTCGCT 83897

#### RESULT 6

Q60985

ID Q60985 standard; DNA; 382 BP.

AC Q60985;

DT 16-MAR-1994 (first entry)

DE Human brain Expressed sequence Tag EST01661.



KM Gene transcription product; genetic markers; tagging; in vivo;  
 KM transcription; mapping; locations; chromosomes; chromosomal, ss.  
 OS Homo sapiens.  
 PN MO9316178-A  
 PD 19-AUG-1993.  
 PF 12-FEB-1993; 001294.  
 PR 12-FEB-1992; US-837195.  
 PR (USSH ) US: DEPT HEALTH & HUMAN SERVICE.  
 PI Adams MD, Moreno RF, Venter CJ;  
 DR WPI: 93-272882/34.  
 PT Enriched oligonucleotides and corresp. sequences - used as  
 PT markers for human genes transcribed in-vivo. facilitate tagging  
 PT of most human genes  
 PS Example 4: Page 421-422; 500pp; English.  
 CC The Expressed Sequence Tag was isolated from a human brain CDNA  
 CC library as part of a large set of ESTs which can be used as markers  
 CC for human genes transcribed in vivo. They can be used to facilitate  
 CC tagging of most human genes, for mapping locations of expressed genes  
 CC on chromosomes, for individual or forensic identification, for mapping  
 CC locations of disease-associated genes, for identification of tissue  
 CC type, and for prepn. of antisense sequences, probes and constructs.  
 CC See also Q59041-061440.  
 SO Sequence 382 BP; 78 A; 97 C; 90 G; 115 T;

[illegible]



KEYWORDS  
SOURCE  
ORGANISM

HTG.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutharia; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

1 (bases 1 to 32070)  
Bentley, D. and Magsi, L.  
The sequence of H. sapiens cosmid LUCA22  
Unpublished (1996)  
2 (bases 1 to 32070)  
Waterston, R.

## REFERENCE

Direct Submission  
Submitted (02-OCT-1996)

## COMMENT

Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA, and  
e-mail: Sapiens@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest

## SOURCE INFORMATION:

This clone is from a chromosome 3 specific library, VECTOR: pWE15  
Clone reference:Wing-Hui Wei et al, CANCER RESEARCH,  
56,1487-1492,1996.

## NEIGHBORING SEQUENCE INFORMATION:

FEATURES  
The left clone is LUCA21: the right clone is 398P81. Actual start  
of this cosmid is at base position 1 of HMDUCA22.  
location/Qualifiers

1. .32070

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="LUCA22"

/clone.lib="LIML3"

/map="3p21.3"

1. .236

/rpt\_family="ALU"

728 .1009

/rpt\_family="ALU"

complement(1876. .2168)

/rpt\_family="ALU"

complement(10132. .10417)

/rpt\_family="ALU"

10458. .10603

/rpt\_family="ALU"

complement(10601. .10786)

/rpt\_family="ALU"

complement(11435. .29883)

/gene="LUCA15"

complement(11435. .11560,12513. .12642,12721. .12818,

13912. .13986,14285. .14464,15641. .15712,15793. .15942,

16348. .16509,19112. .19200,19419. .19503,20194. .20279,

21578. .21650,21735. .21812,22312. .22398,23017. .23114,

24169. .24329,24736. .24801,25570. .25630,26712. .26795,

29269. .29342,29823. .29883))

/gene="LUCA15"

/note="H\_LUCA22.1: coded for by human CDNA5 U23946,

N41616, D19618, N23866, T03540, R45218, T10051, N29616,  
T10050, W76177, R21564, U23946, R24955 and T99027"

/codon\_start=1

/product="partial CDS, human putative tumor suppressor

(U23946)"

/protein\_id="AA042216.1"

/db\_xref="GI:1613900"

/translation="MSESSEGPQADYRLMKRTGVCSEFAVEFHHQDQTSMEAN

QKRLVQGNHIMHSNPRKEEDWLCNKKLNRRKRLKRCRGADKFSDEQEPFG

TTESVSVDYICDITLTKNLIAPHTVDSIMTALSPASIVANNRLIKDQTKNGF

AFVQLSSADASQLQIILQSLHPPKIDCKTICVDPKASARNDVLSDGNSVAFSA

STALAAQWSSSTOSQSGSVSYLQPDQDQAVAYQVSDYQDFYQQQAGLSD

ASSASCTAVTTTSAVVSQSPOLYNTPSPGPEEQSTSTQAPKASFTGVAP

GTRKVPDSTQYDESSGYYDPDTGLYDPSNOYVNSLQOYLNDGKEFTVPAA

ESSHQOSCLPPAKGKEKPKSKTAQOILAKDMERNKSLNKQKERNFQVNS

LREERRESAADAGFALFEKGAARQQLPELVKNGDEBNPLKGLVAAYSGSD

NBELVERLESEERKADMKKACLCRRQFPKDALYHOOOLSDLRKNDYRRSR

LSQELFALEREREMKRDRAARRKYGIPPEPPKROFQAGTVNYEPTKNGI

DHSNIGKMLQMGMRGSGGRKQCGITPPIEAQVRLKAGLGAASVGLGADSY

KDAVRKAMRAREPME"

complement(12106. .12379)

/rpt\_family="ALU"

complement(14318. .14464)

/gene="LUCA15"

/note="Grail prediction, score = 88"

repeat\_region

complement(14715. .15245)

/rpt\_family="ALU"

complement(15602. .15712)

/gene="LUCA15"

/note="Grail prediction, score = 82"

16400. .16739

/gene="LUCA15"

/note="match: human 3' EST N74644 bases 1-340"

16400. .16673

/gene="LUCA15"

/note="16739"

16673. .16739

/gene="LUCA15"

/note="match: human 5' EST W00982 bases 283-343"

16740. .17021

/rpt\_family="ALU"

complement(17457. .17759)

/rpt\_family="ALU"

18209. .18449

/rpt\_family="ALU"

18664. .18737

/gene="LUCA15"

/note="match: human 3' EST NS9570 bases 355-428"

complement(19663. .19951)

/rpt\_family="ALU"

20510. .20800

/rpt\_family="ALU"

complement(21166. .21455)

/rpt\_family="ALU"

21898. .22188

/rpt\_family="ALU"

23353. .23915

/rpt\_family="ALU"

complement(24579. .24684)

/rpt\_family="ALU"

24685. .25296

/gene="LUCA15"

/note="close match: multiple ESTs, see AA001762 and

R01204"

complement(25227. .25519)

/rpt\_family="ALU"

complement(26223. .26512)

/rpt\_family="ALU"

27411. .27822

/gene="LUCA15"

/note="close match: human 3' EST AA037589 and 5' EST

AA03758"

28604. .28733

misc\_feature

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:50:51 ; Search time 9481.22 Seconds

(without alignments)  
-861.153 Million cell updates/sec

Title: US-09-090-672B-2

Perfect score: 2689

Sequence: 1 GTTGAGGCTCTGGGGCGCA.....AAAAAAAAAAAAAAAAAAAA 2689

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched:

821193 seqs, -1518192014 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database:

GenBank:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pi1:\*  
8: gb\_pi2:\*  
9: gb\_pi3:\*  
10: gb\_pi4:\*  
11: gb\_pi5:\*  
12: gb\_pi6:\*  
13: gb\_pi7:\*  
14: gb\_pi8:\*  
15: gb\_pi9:\*  
16: gb\_pi10:\*  
17: gb\_pi11:\*  
18: gb\_pi12:\*  
19: gb\_pi13:\*  
20: gb\_pi14:\*  
21: gb\_pi15:\*  
22: gb\_pi16:\*  
23: gb\_pi17:\*  
24: gb\_pi18:\*  
25: gb\_pi19:\*  
26: gb\_pi20:\*  
27: gb\_pi21:\*  
28: gb\_pi22:\*  
29: gb\_pi23:\*  
30: gb\_pi24:\*  
31: gb\_pi25:\*  
32: gb\_pi26:\*  
33: gb\_pi27:\*  
34: gb\_pi28:\*  
35: gb\_pi29:\*  
36: gb\_pi30:\*  
37: gb\_pi31:\*  
38: gb\_pi32:\*  
39: gb\_pi33:\*  
40: gb\_pi34:\*  
41: gb\_pi35:\*  
42: gb\_pi36:\*  
43: gb\_pi37:\*

44: gb\_pi38:\*  
45: gb\_pi39:\*  
46: gb\_pi40:\*  
47: gb\_pi41:\*  
48: gb\_pi42:\*  
49: gb\_pi43:\*  
50: gb\_pi44:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	57.5	32070	10 HUMLUCA22	U73168 Human cosmid
2	518	19.3	3135	40 AF091263	AF091263 Homo sapi
3	467	17.4	3094	40 AF103802	AF103802 Homo sapi
4	379	14.1	2575	10 HSU23946	U23946 Human putat
5	168	6.2	83030	40 AC004814	AC004814 Homo sapi
6	120	4.5	132	5 A74798	A74798 Sequence 48
7	120	4.5	132	5 A74798	A74798 Sequence 48
8	120	4.5	132	5 A74798	A74798 Sequence 48
9	120	4.5	132	5 A74798	A74798 Sequence 48
10	120	4.5	132	5 A74798	A74798 Sequence 48
11	120	4.5	132	5 A74798	A74798 Sequence 48
12	120	4.5	132	5 A74798	A74798 Sequence 48
13	120	4.5	132	5 A74798	A74798 Sequence 48
14	120	4.5	132	5 A74798	A74798 Sequence 48
15	120	4.5	132	5 A74798	A74798 Sequence 48
16	120	4.5	132	5 A74798	A74798 Sequence 48
17	120	4.5	132	5 A74798	A74798 Sequence 48
18	120	4.5	132	5 A74798	A74798 Sequence 48
19	120	4.5	132	5 A74798	A74798 Sequence 48
20	120	4.5	132	5 A74798	A74798 Sequence 48
21	120	4.5	132	5 A74798	A74798 Sequence 48
22	120	4.5	132	5 A74798	A74798 Sequence 48
23	120	4.5	132	5 A74798	A74798 Sequence 48
24	120	4.5	132	5 A74798	A74798 Sequence 48
25	120	4.5	132	5 A74798	A74798 Sequence 48
26	120	4.5	132	5 A74798	A74798 Sequence 48
27	120	4.5	132	5 A74798	A74798 Sequence 48
28	120	4.5	132	5 A74798	A74798 Sequence 48
29	120	4.5	132	5 A74798	A74798 Sequence 48
30	120	4.5	132	5 A74798	A74798 Sequence 48
31	120	4.5	132	5 A74798	A74798 Sequence 48
32	120	4.5	132	5 A74798	A74798 Sequence 48
33	120	4.5	132	5 A74798	A74798 Sequence 48
34	120	4.5	132	5 A74798	A74798 Sequence 48
35	120	4.5	132	5 A74798	A74798 Sequence 48
36	120	4.5	132	5 A74798	A74798 Sequence 48
37	120	4.5	132	5 A74798	A74798 Sequence 48
38	120	4.5	132	5 A74798	A74798 Sequence 48
39	120	4.5	132	5 A74798	A74798 Sequence 48
40	120	4.5	132	5 A74798	A74798 Sequence 48
41	120	4.5	132	5 A74798	A74798 Sequence 48
42	120	4.5	132	5 A74798	A74798 Sequence 48
43	120	4.5	132	5 A74798	A74798 Sequence 48

#### ALIGNMENTS

RESULT 1  
LOCUS HUMLUCA22/c  
DEFINITION Human cosmid LUCA22, complete sequence.  
ACCESSION U73168  
VERSION U73168.1 GI:1613899

07-FEB-1997

/gene="LUCA15"		/note="close match; human EST Z20369 bases 1-129"	
misc_feature		29212..29547	
repeat_region		/gene="LUCA15"	
repeat_region		/note="close match; rat 3' EST H33623"	
repeat_region		complement(29990..30410)	
repeat_region		/rpt_family="ALU"	
repeat_region		complement(30609..30901)	
repeat_region		/rpt_family="ALU"	
repeat_region		30996..31595	
repeat_region		/rpt_family="ALU"	
repeat_region		complement(31762..32048)	
repeat_region		/rpt_family="ALU"	
BASE COUNT		8573 a 8148 c 7445 g 7904 t	
ORIGIN			
Query Match		57.5% Score 1547; DB 10; Length 32070;	
Best Local Similarity		99.6% Pred. No. 0;	
Matches 2197; Conservative		0; Mismatches 1; Indels 7; Gaps	
QY	446	ATTTCGAGAATGATGGAGCTCTCGAAGGCCCTCAGCCTCGGATGGAGCTGATGAAG	505
Db	29892	ATTTCGAGAATGATGGAGCTCTCGAAGGCCCTCAGCCTCGGATGGAGCTGATGAAG	29833
QY	506	AGAAACAGGTGAGAGCTTCTTAGTTCCTGATATATTATTGTTCTCTCCCATTCGCCAC	565
Db	29832	AGAAACAGGTGAGAGCTTCTTAGTTCCTGATATATTATTGTTCTCTCCCATTCGCCAC	29773
QY	566	CTCAGTCCCTAAAGAACATCTCGATTCCCCAGCTTCTCAAGCACATGAATTCAGAAATGAA	625
Db	29772	CTCAGTCCCTAAAGAACATCTCGATTCCCCAGCTTCTCAAGCACATGAATTCAGAAATGAA	29713
QY	626	AGGTTTGCCATGGCTTAAGGAATGTGACTCTTTGAAACCATGTTAGCATCTGAGGAATTT	685
Db	29712	AGGTTTGCCATGGCTTAAGGAATGTGACTCTTTGAAACCATGTTAGCATCTGAGGAATTT	29653
QY	686	TTTTAAACTTTGTTTAGGACTTTTTTTTCCCTTAGTAACTAATGATTTATAAACTCCT	745
Db	29652	TTTTAAACTTTGTTTAGGACTTTTTTTTCCCTTAGTAACTAATGATTTATAAACTCCT	29593
QY	746	TTTTTTTTTTTGTACTATAGTCGGTTGCATGTTTACTTTAAGCGTGGAAATCAAAATGGAGT	805
Db	29593	TTTTTTTTTTTGTACTATAGTCGGTTGCATGTTTACTTTAAGCGTGGAAATCAAAATGGAGT	29535
QY	806	GGCATTTAGTTCAGCGGCTGTTTCTTTCGCAATGGCAAGTATCAAGAAGATCCCCAAGT	865
Db	29534	GGCATTTAGTTCAGCGGCTGTTTCTTTCGCAATGGCAAGTATCAAGAAGATCCCCAAGT	29475
QY	866	CAAGTCACATTTGTAAGCTGCTCCCAATTTGGCTTTGTCACGCAGTGTGAAGCAGTGG	925
Db	29474	CAAGTCACATTTGTAAGCTGCTCCCAATTTGGCTTTGTCACGCAGTGTGAAGCAGTGG	29415
QY	926	GAGAGAGATTCACCTGTTTATAAGGAACGTGACTAACACAAGTATCCCGTCTATATCTGAA	985
Db	29414	GAGAGAGATTCACCTGTTTATAAGGAACGTGACTAACACAAGTATCCCGTCTATATCTGAA	29355
QY	986	TGCTGTCTCTAGTGTGAAGCCGTGGTTTTCGCTTTCGTGGAGTTTATCAGTTGCAAGATG	1045
Db	29354	TGCTGTCTCTAGTGTGAAGCCGTGGTTTTCGCTTTCGTGGAGTTTATCAGTTGCAAGATG	29295
QY	1046	CTACCAGCTGGATGGAAGCCATCAGTTGCTTCACTACCAAGTCTAGATATTCATGAA	1105
Db	29294	CTACCAGCTGGATGGAAGCCATCAGTTGCTTCACTACCAAGTCTAGATATTCATGAA	29235
QY	1106	AATGGAACAGTCTGTACAAATTTTAAAAAAGGTTGAAGAGTGGTTGTTCCAAAGGAG	1165
Db	29234	AATGGAACAGTCTGTACAAATTTTAAAAAAGGTTGAAGAGTGGTTGTTCCAAAGGAG	29175
QY	1166	TGACTTTTTTTTAAAAAAGGCTTTCTATATATATAAATTTGATGTTACTAGATAAGT	1225
Db	29174	TGACTTTTTTTTAAAAAAGGCTTTCTATATATATAAATTTGATGTTACTAGATAAGT	29116

QY	1226	ACAGTACCAGGACCTTCATTATAGAAATTTGTTCTGCCCTTTAAACATGGCTACCTACCTGG	1285
DB	29115	ACAGTACCAGGACCTTCATTATAGAAATTTGTTCTGCCCTTTAAACATGGCTACCTACCTGG	29056
QY	1286	CAGGGCTTTGTTAACTACTGAATACCTGCTCGTAAATCAGTAAACATCTTAATGTTTCC	1345
DB	29055	CAGGGCTTTGTTAACTACTGAATACCTGCTCGTAAATCAGTAAACATCTTAATGTTTCC	28996
QY	1346	CTTTTTCTTAGTTTGTATATATCTTATTCCTATTATGCCATTGAGAGTAAGCTTAGTATACAA	1405
DB	28995	CTTTTTCTTAGTTTGTATATATCTTATTCCTATTATGCCATTGAGAGTAAGCTTAGTATACAA	28936
QY	1406	CTCTCCATTGACAGTGAAGAGACATAGTGAAGTCTGTGGCGGCATTTTATAGTAA	1465
DB	28935	CTCTCCATTGACAGTGAAGAGACATAGTGAAGTCTGTGGCGGCATTTTATAGTAA	28876
QY	1466	TTCCCTTTTCTGCCCTGAAGACACCAAAAGCCCTCTGGAGCGCTAACTGCTCAGACCGGTC	1525
DB	28875	TTCCCTTTTCTGCCCTGAAGACACCAAAAGCCCTCTGGAGCGCTAACTGCTCAGACCGGTC	28816
QY	1526	TTCAGGGAATATTTAAGGACTTAGTGGAAATTTATGAACAATAAGTCTGATGAGATTAGCC	1585
DB	28815	TTCAGGGAATATTTAAGGACTTAGTGGAAATTTATGAACAATAAGTCTGATGAGATTAGCC	28756
QY	1586	TGGAGTGTGTCTCTCAGCTGTCTAATCTAGTTAGAGTGGCATTAAACATTCTAATCTCC	1645
DB	28755	TGGAGTGTGTCTCTCAGCTGTCTAATC---TAGAGTGGCATTAAACATTCTAATCTCC	28700
QY	1646	TTGAGAAATGCCCTTTATAGTCTGTCAAAAGCAAGTCATGATGTTCTTCGAGGTAGTGT	1705
DB	28699	TTGAGAAATGCCCTTTATAGTCTGTCAAAAGCAAGTCATGATGTTCTTCGAGGTAGTGT	28640
QY	1706	TAACTGAAAGTGTCTTCTCAGTTTGTCAAGATAATGTTTCTAGTGTGGCATTAAATAACAT	1765
DB	28639	TAACTGAAAGTGTCTTCTCAGTTTGTCAAGATAATGTTTCTAGTGTGGCATTAAATAACAT	28580
QY	1766	TTTTTGCAGAACTCCAAAGCGACATTTTGAATGCCCTTTAAACAAGTCATTCTCGGAG	1825
DB	28579	TTTTTGCAGAACTCCAAAGCGACATTTTGAATGCCCTTTAAACAAGTCATTCTCGGAG	28520
QY	1826	TTTGTCTTGACTCATTTATCTGCTTTTCTCGCAGCATTTCTGATTTGAGTCATCCATGAAT	1885
DB	28519	TTTGTCTTGACTCATTTATCTGCTTTTCTCGCAGCATTTCTGATTTGAGTCATCCATGAAT	28460
QY	1886	CCATGAATAAAGTTACATTTTGTATGTTGATATTTGCCATTTATAACAAGACTCAGTA	1945
DB	28459	CCATGAATAAAGTTACATTTTGTATGTTGATATTTGCCATTTATAACAAGACTCAGTA	28400
QY	1946	ATCAGGATATCATTGTTGACTGACTGATTTGTTAAAGTTTTTAAAGCCTCTCATTTTCCCTAA	2005
DB	28399	ATCAGGATATCATTGTTGACTGACTGATTTGTTAAAGTTTTTAAAGCCTCTCATTTTCCCTAA	28340
QY	2006	CCAGAAATCAGACCTGATTTTATTTAAAGTAGAGCTTCATTCATTCATACCATAGAT	2065
DB	28339	CCAGAAATCAGACCTGATTTTATTTAAAGTAGAGCTTCATTCATTCATACCATAGAT	28280
QY	2066	ACCATCTCTAGTAATCCAGAACATATACAAGGTTTCATGTGAGTCTGCTTTCTTGACATGA	2125
DB	28279	ACCATCTCTAGTAATCCAGAACATATACAAGGTTTCATGTGAGTCTGCTTTCTTGACATGA	28220
QY	2126	TAGCATTTGTTGATGTCAGTGGATATGTCAGAAATGACTAACCTTAGGAGTTTAAACCTCCTA	2185
DB	28219	TAGCATTTGTTGATGTCAGTGGATATGTCAGAAATGACTAACCTTAGGAGTTTAAACCTCCTA	28160
QY	2186	AGAACTAAAACCTGTAAGACATTTAAAAGTCTCCACATTTTAAATGATACAAAGCTAT	2245
DB	28159	AGAACTAAAACCTGTAAGACATTTAAAAGTCTCCACATTTTAAATGATACAAAGCTAT	28100
QY	2246	GTTACTGTGTAACACATTTACAGTTCAAATTCACCTCCAGAAATAAAAGGCCAGTAGGATTA	2305
DB	28099	GTTACTGTGTAACACATTTACAGTTCAAATTCACCTCCAGAAATAAAAGGCCAGTAGGATTA	28040
QY	2306	GGGACTCACTGGTAGTTTGGAGTCTCCACACACATCCCTCCTCTGTGGGATGATCTATT	2365

```
|||||
Db 28039 GGGACTCACTGGTGGAGTCTCCAGACACATCCCTCTAGTGGATGATCTATT 27980
Qy 2366 CACATATCTCCAGCTTTTATTTTCTCTGTATATACAGTGAAGTGGCCCTT 2425
|||||
Db 27979 CACATATCTCCAGCTTTTATTTTCTCTGTATATACAGTGAAGTGGCCCTT 27920
Qy 2426 CAGCTTTTCTCTCTGCGCAGACATGAGTCTTGCCTTTAGATATGCGAGACAAAAT 2485
|||||
Db 27919 CAGCTTTTCTCTCTGCGCAGACATGAGTCTTGCCTTTAGATATGCGAGACAAAAT 27860
Qy 2486 TCACAGATGCTTAAATCTTCAGGATTCAGACACAAAATGCTACAGTATGAT 2545
Db 27859 TCACAGATGCTTAAATCTTCAGGATTCAGACACAAAATGCTACAGTATGAT 27800
Qy 2546 GTTTAGAGGGTTAGACTCTCTTTTAAATCTGGATATCTAACACCTACTTAAATCTGT 2605
|||||
Db 27799 GTTTAGAGGGTTAGACTCTCTTTTAAATCTGGATATCTAACACCTACTTAAATCTGT 27740
Qy 2606 TTGATAGTGTCAAAACCCACCCCTTGATCTTCCACCCCA 2650
|||||
Db 27739 TTGATAGTGTCAAAACCCACCCCTTGATCTTCCACCCCA 27695

RESULT 2
AF091263
LOCUS AF091263 3135 bp mRNA PRI 06-OCT-1999
DEFINITION Homo sapiens RNA binding motif protein 5 (RBM5) mRNA, complete cds.
ACCESSION AF091263
VERSION AF091263.1 GI:4140646
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Eist,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
A evolutionary rearrangement of the Xp11.3-11.23 region in 3p21.3,
a region frequently deleted in a variety of cancers
Genomics 60 (2), 238-240 (1999)
99417685
MEDLINE
REFERENCE
2 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Eist,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
A comparison of genomic structures and expression patterns of two
closely related flanking genes in a critical lung cancer region at
3p21.3
Unpublished
JOURNAL
3 (bases 1 to 3135)
Timmer,T., Terpstra,P., van den Berg,A., Veldhuis,P.M.J.F., Ter
Eist,A., Voutsinas,M., Hulsbeek,M.M.F., Draaijers,T.G.,
Looman,M.W.G., Kok,K., Naylor,S.L. and Buys,C.H.C.M.
Direct Submission
JOURNAL
Submitted (13-SEP-1998) Medical Genetics, University Groningen,
A.Deusinglaan 4, Groningen 9713AW, The Netherlands
FEATURES
Location/Qualifiers
1..3135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/notes="lung cancer region"
1..3135
/gene="RBM5"
149..2596
/gene="RBM5"
/codon_start=1
/product="RNA binding motif protein 5"
/protein_id="A0D04159.1"
/db_xref="GI:4140647"
/translation="MGSDFRVSTERSGRYGIIDRRDRERSRRRDSYDKRSSD
```

```
DRGRDYDDYRDYDSPERERERNRSDSDYHSDGYGEHDIRHDSIDRESKTIIML
RGLPITTESIDIREMESFEQPOADVRLMKRKTVSGRFAVEFYHLDQATSWMEAN
QKLIQKHYAMHYSNRPKFDWLCNCKLNFRKLRKFCRCGADKFDSEQVPPG
TTESVQSDYVCOTILRLNAPHTVDSIMTASPYASLAVANNIRLIKQIQOQNRGF
AFVOLSNDASQLLOLQSLHPPKIDKTIQVDFAKSARDLVLSNRRVSFAFSVA
STATAAOWSSTQSGEGSDYSYLPQGDGYAQYQYSDYQDYFOOQAGGLES
ASSAGTAVTTTAAVVSQSPLYNOSNPPSPTEEAQPSSTSTQAPAASTGYVP
GTYAVPDTSTYQYDESSGYIDPTTGLYDPSQYYSNLTQYLYWDEKETYVPA
AESSHOOSGLPPAKGKKEKPKSKTAQIADKMERWAKSLNKKOKFNKNSFOVFN
DNEELERESNAADAGFALFEKKGALAEQOLPELVNRCDEENPLKGLVAAYSDDS
RLSEQUELEALEREMKYRDAAERREKIGDPEPPKRRKKQFDAGTVNYDPTKDG
YKDAVRKAMFAREFTE"
566..589
misc_feature /gene="RBM5"
/notes="RNA-binding motif; unclassified site"
707..760
misc_feature /gene="RBM5"
/notes="zinc finger (C(x2)C(x10)C(x2)C) motif; DNA binding
site"
968..991
misc_feature /gene="RBM5"
/notes="RNA-binding motif; unclassified site"
2093..2164
misc_feature /gene="RBM5"
/notes="zinc finger (C(x2)C(x12)H(x6)H) motif; DNA binding
site"
2270..2320
misc_feature /gene="RBM5"
/notes="bipartite nuclear signal motif; signal-peptide
site"
BASE COUNT 900 a 705 c 822 g 708 t
ORIGIN
Query Match 19.3%; Score 518; DB 40; Length 3135;
Best Local Similarity 100.0%; Pred. No. 3.5e-276;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTTGAGGCTTCTGGGCGCAGAACCGCTACTGCTGCTCGGTCCTCTCTCTGGGAAAAAT 60
Db 43 GTTGAGGCTTCTGGGCGCAGAACCGCTACTGCTGCTCGGTCCTCTCTCTGGGAAAAAT 102
Qy 61 AAAATTTCAACCTTTTGAGCTGTCTCTAAATCTTCAGTGGGACAAATGGGTTCAGACAA 120
Db 103 AAAATTTCAACCTTTTGAGCTGTCTCTAAATCTTCAGTGGGACAAATGGGTTCAGACAA 162
Qy 121 AAGAGTGAAGTGAACAGAGCGCTAGTGAAGATACGGTTCCATCATAGACAGGATGACCG 180
Db 163 AAGAGTGAAGTGAACAGAGCGCTAGTGAAGATACGGTTCCATCATAGACAGGATGACCG 222
Qy 181 TGATGAGCGTGAATCCCAAGCAGCGGAGGAGCTCAGATTACAAAAGATCTAGTGATGA 240
Db 223 TGATGAGCGTGAATCCCAAGCAGCGGAGGAGCTCAGATTACAAAAGATCTAGTGATGA 282
Qy 241 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATGACAGTCCAGAGAGAGAGCGTGA 300
Db 283 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATGACAGTCCAGAGAGAGAGCGTGA 342
Qy 301 AAGAGGAACAGTACCGGATCCGAAGATGCGTACCATTCAGATGGTGCATGGTGAACA 360
Db 343 AAGAGGAACAGTACCGGATCCGAAGATGCGTACCATTCAGATGGTGCATGGTGAACA 402
Qy 361 CGACTATAGGCATGACATCAGTGAAGAGGAGGAGCAAGACCATCATGCTGCGCGGCT 420
Db 403 CGACTATAGGCATGACATCAGTGAAGAGGAGGAGCAAGACCATCATGCTGCGCGGCT 462
Qy 421 TCCCATCACCATCAGAGAGCGGATATTCGAGAAATGATGAGTCTCTTCGAGGCCCTCA 480
Db 463 TCCCATCACCATCAGAGAGCGGATATTCGAGAAATGATGAGTCTCTTCGAGGCCCTCA 522
Qy 481 GCCTCGGATGTGAGGCTGATGAAGAGAGAAACAGGTG 518
```



```
|||||
Db 523 GCCTGGGATGTGAGCTGATGAGAGGAACAGGTG 560
|||||
RESULT 3
AF103802
LOCUS Homo sapiens clone H37 unknown mRNA. PRI 16-OCT-1999
DEFINITION AF103802
ACCESSION AF103802.1 GI:6048971
VERSION FLI_CDNA.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3094)
AUTHORS Oh,J.J., Grosshans,D.R., Wong,S.G. and Slamon,D.J.
TITLE Identification of differentially expressed genes associated with
HER-2/neu overexpression in human breast cancer cells
JOURNAL Nucleic Acids Res. 27 (20), 4008-4017 (1999)
MEDLINE 99428654
REFERENCE 2 (bases 1 to 3094)
AUTHORS Oh,J.J., Wong,S.G. and Slamon,D.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) Hematology/Oncology, UCLA School of
Medicine, 5-535 MRL, 675 Circle Drive South, Los Angeles, CA 90095,
USA
FEATURES
source
Location/Qualifiers
1..3094
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="H37"
/cell_line="MCF-7/HER2"
/cell_type="breast adenocarcinoma"
/note="MCF-7 cells retrovirally transformed with human
HER-2/neu full-length cDNA"
127..2574
/note="expression increases in HER-2/neu overexpressing
cells; differentially expressed."
/codon_start=1
/product="unknown"
/protein_id="AA02422.1"
/db_xref="GI:6048972"
/translation="MGSDRVRSTERSGRYGSIIIDRRDRERSRRRSDYKRSSD
DRGDRYDRYDPSRPERRRNSDRSDGYHSDGYGHDYRHDISDERESKTM
RGLPITITSDIREMMESFEGQPADVRLMKRTGVRGFAPVFYHLQDATSWMEAN
QKLVIGKHIAHMYGNRPKFEKMLCNKCLNFRKRLKFCGADKFDSEQVPPG
TTESQVSDYICDTIILRNIAPIHTVVDISMTALSPYASLVNINRLIKDKQTQNRGF
AFVQLSAMDASOLLQIOLHPLKIDGKTIGDVFPAKSARKDLVSDGNRVSAFVA
STATAAROSSTOSQSGGSDYVSYLOPGDGYAQYQYQDYQOYQOAGLESD
ASSASTAVITTSAAVVSQPOLYQTSNPPSPSTEAPQSTSTQAPASPTGVTP
GKIYAPDTSYDIDESSGYIDPTGLYDNPNSOYIYNSUTQQLYWDGKEIYVA
AESSHQOGLPAPKPKPKPKPKKTAQQLIAKDMRWAKSLNKKRNFKNFQPVN
SLREPERRESAADAGFALFKKGALEAQQLIPELVRNGDEPNLKRGLVAAYSQDS
DNEEELVERLESSEELKADWKMACLCRRQPNKDALVRHQQLSDLHKQNMIDYRGS
RLSEOLELELEREMKYRDRAAERREKYGIPEPEPKRKQFDAGTVNYEQPTKDG
IDHSNIGNMLOAMWREGSGGLRKGKQGITATIEAQVRLKAGLKAQSGAYGLSGADS
YKDAVRKMPARTEME"
BASE COUNT 879 a 702 c 811 g 701 t 1 others
ORIGIN
Query Match 17.4% Score 467; DB 40; Length 3094;
Best Local Similarity 99.8%; Prod. No. 6.9e-248;
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTTGAGGTTCTGGGGCGACAGCCGCTACTGCTGCTTCGGTCTCTCTTGGGAAAAAT 60
|||||
Db 21 GTTGAGGTTCTGGGGCGACAGCCGCTACTGCTGCTCTCTCTTGGGAAAAAT 80
|||||
Qy 61 AAAATTCAACCTTTTGAGCTGTGCTCTCTCTTCAAGTGGGACAAATGGTTTCACACAA 120
|||||
Db 81 AAAATTCAACCTTTTGAGCTGTGCTCTCTCTTCAAGTGGGACAAATGGTTTCACACAA 140
|||||
QY 121 AAGAGTGAAGTGAAGAGCGGTAGTGAAGATACGGTTCCATCATAGACGGATGACCG 180
|||||
Db 141 AAGAGTGAAGTGAAGAGCGGTAGTGAAGATACGGTTCCATCATAGACGGATGACCG 200
|||||
QY 181 TGATAGCGCTGAATCCCGAAGCAGCGGAGGACTCAGATTACAAAAGATCTAGTGATGA 240
|||||
Db 201 TGATAGCGCTGAATCCCGAAGCAGCGGAGGACTCAGATTACAAAAGATCTAGTGATGA 260
|||||
QY 241 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATCAGAGTCCAGAGAGAGCGGTGA 300
|||||
Db 261 TCGGAGGGGTGATAGATATGATGACTACCGAGACTATCAGAGTCCAGAGAGAGCGGTGA 320
|||||
QY 301 AAGAAGAACAGTGAACCGATCCGAAGATGGTTCACATTTCAGATGGTGAAGCA 360
|||||
Db 321 AAGAAGAACAGTGAACCGATCCGAAGATGGTTCACATTTCAGATGGTGAAGCA 380
|||||
QY 361 CGACTATAGCGATGACATCAGTGACGAGAGGAGAGCAAGACCATCATGCTCGCGGCGCT 420
|||||
Db 381 CGACTATAGCGATGACATCAGTGACGAGAGGAGAGCAAGACCATCATGCTCGCGGCGCT 440
|||||
QY 421 TCCCATCACCATCACAGAGCGGATATTCGAGAAATGATGGAGCTCTCGAAGCGCCTCA 480
|||||
Db 441 TCCCATCACCATCACAGAGCGGATATTCGAGAAATGATGGAGCTCTCGAAGCGCCTCA 500
|||||
QY 481 GCCTGGGATGTGAGGCTGATCAAGAGGAGAAACAGGTG 518
|||||
Db 501 GCCTGGGATGTGAGGCTGATCAAGAGGAGAAACAGGTG 538
|||||
RESULT 4
HSU23946
LOCUS Human putative tumor suppressor (LUCA15) mRNA. PRI 16-MAY-1996
DEFINITION U23946
ACCESSION U23946
VERSION U23946.1 GI:1244403
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2575)
AUTHORS Bader,S., Latif,F., Duh,F., Wei,M., Kashuba,V., Sekido,Y., Lee,C.,
Koonin,E., Zabarofsky,E., Klein,G., Minna,J.D. and Lerman,M.
TITLE A putative tumor suppressor gene LUCA15 on 3p21.3 encodes two RNA
recognizing motifs and is related to the Drosophila tumor
suppresorgene Sxl
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2575)
AUTHORS Duh,F.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-1995) Fuh-Mei Duh, BCDP, PRI/ Dyncorp, NCI-FCRDC,
Building 560, Rm 12-71, Frederick, MD 21702, USA
FEATURES
source
Location/Qualifiers
1..2575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LUCA15"
/chromosome="3"
/map="3p21.3"
/cell_type="islet beta-cells"
1..69
70..2517
/gene="LUCA15"
70..2517
/gene="LUCA15"
/note="putative tumor suppressor"
/codon_start=1
/protein_id="AAA99715.1"
/db_xref="GI:1244404"
/translation="MGSDKRVSRSTERSGRYGSIIIDRRDRERSRRRSDYKRSSD
DRGDRYDGSRDYDPSRPERRRNSDRSDGYHSDGYGHDYRHDISDERESKTM
RGLPITITSDIREMMESFEGQPADVRLMKRTGVRGFAPVFYHLQDATSWMEAN
QKLVIGKHIAHMYGNRPKFEKMLCNKCLNFRKRLKFCGADKFDSEQVPPG
TTESQVSDYICDTIILRNIAPIHTVVDISMTALSPYASLVNINRLIKDKQTQNRGF
AFVQLSAMDASOLLQIOLHPLKIDGKTIGDVFPAKSARKDLVSDGNRVSAFVA
STATAAROSSTOSQSGGSDYVSYLOPGDGYAQYQYQDYQOYQOAGLESD
ASSASTAVITTSAAVVSQPOLYQTSNPPSPSTEAPQSTSTQAPASPTGVTP
GKIYAPDTSYDIDESSGYIDPTGLYDNPNSOYIYNSUTQQLYWDGKEIYVA
AESSHQOGLPAPKPKPKPKKTAQQLIAKDMRWAKSLNKKRNFKNFQPVN
SLREPERRESAADAGFALFKKGALEAQQLIPELVRNGDEPNLKRGLVAAYSQDS
DNEEELVERLESSEELKADWKMACLCRRQPNKDALVRHQQLSDLHKQNMIDYRGS
RLSEOLELELEREMKYRDRAAERREKYGIPEPEPKRKQFDAGTVNYEQPTKDG
IDHSNIGNMLOAMWREGSGGLRKGKQGITATIEAQVRLKAGLKAQSGAYGLSGADS
YKDAVRKMPARTEME"
5'UTR
gene
1..69
70..2517
/gene="LUCA15"
70..2517
/gene="LUCA15"
/note="putative tumor suppressor"
/codon_start=1
/protein_id="AAA99715.1"
/db_xref="GI:1244404"
/translation="MGSDKRVSRSTERSGRYGSIIIDRRDRERSRRRSDYKRSSD
DRGDRYDGSRDYDPSRPERRRNSDRSDGYHSDGYGHDYRHDISDERESKTM
RGLPITITSDIREMMESFEGQPADVRLMKRTGVRGFAPVFYHLQDATSWMEAN
QKLVIGKHIAHMYGNRPKFEKMLCNKCLNFRKRLKFCGADKFDSEQVPPG
TTESQVSDYICDTIILRNIAPIHTVVDISMTALSPYASLVNINRLIKDKQTQNRGF
AFVQLSAMDASOLLQIOLHPLKIDGKTIGDVFPAKSARKDLVSDGNRVSAFVA
STATAAROSSTOSQSGGSDYVSYLOPGDGYAQYQYQDYQOYQOAGLESD
ASSASTAVITTSAAVVSQPOLYQTSNPPSPSTEAPQSTSTQAPASPTGVTP
GKIYAPDTSYDIDESSGYIDPTGLYDNPNSOYIYNSUTQQLYWDGKEIYVA
AESSHQOGLPAPKPKPKPKKTAQQLIAKDMRWAKSLNKKRNFKNFQPVN
SLREPERRESAADAGFALFKKGALEAQQLIPELVRNGDEPNLKRGLVAAYSQDS
DNEEELVERLESSEELKADWKMACLCRRQPNKDALVRHQQLSDLHKQNMIDYRGS
RLSEOLELELEREMKYRDRAAERREKYGIPEPEPKRKQFDAGTVNYEQPTKDG
IDHSNIGNMLOAMWREGSGGLRKGKQGITATIEAQVRLKAGLKAQSGAYGLSGADS
YKDAVRKMPARTEME"
BASE COUNT 879 a 702 c 811 g 701 t 1 others
ORIGIN
Query Match 17.4% Score 467; DB 40; Length 3094;
Best Local Similarity 99.8%; Prod. No. 6.9e-248;
Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTTGAGGTTCTGGGGCGACAGCCGCTACTGCTGCTCTCTTGGGAAAAAT 60
|||||
Db 21 GTTGAGGTTCTGGGGCGACAGCCGCTACTGCTGCTCTCTTGGGAAAAAT 80
|||||
Qy 61 AAAATTCAACCTTTTGAGCTGTGCTCTCTCTTCAAGTGGGACAAATGGTTTCACACAA 120
|||||
Db 81 AAAATTCAACCTTTTGAGCTGTGCTCTCTCTTCAAGTGGGACAAATGGTTTCACACAA 140
|||||
```



Db 60 ATTGATGGTCTTCGAGGTAGTCTTAACCTGAAGTGTCTTCTCAGTTGTTCAGATAATGTT 1

RESULT 7  
LOCUS A77777/c 132 bp DNA PAT 19-OCT-1999  
DEFINITION Sequence 484 from Patent EP0587279.  
ACCESSION A77777  
VERSION A77777.1 GI:6089442  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 132)  
AUTHORS Slobin,D.R. and Hadfield,K.M.  
TITLE HUMAN NUCLEIC ACID FRAGMENTS ISOLATED FROM BRAIN, ADRENAL TISSUE,  
PLACENTA OR BONE MARROW AND THEIR USE  
JOURNAL Patent: EP 0587279-A 16-MAR-1994;  
MEDICAL RES COUNCIL (GB)  
FEATURES Location/Qualifiers  
source 1. .132  
/organism="unidentified"  
/db\_xref="taxon:32644" 37 t

BASE COUNT 49 a 26 c 20 g 37 t  
ORIGIN

Query Match 4.5%; Score 120; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 2.3e-55;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1622 AGTGCATTACATCTTAATCTCCTTGAGAAATGCTTTTATAGTCTGTTCAAGCAAGTC 1681  
|||||  
Db 120 AGTGCATTACATCTTAATCTCCTTGAGAAATGCTTTTATAGTCTGTTCAAGCAAGTC 61  
|||||

Qy 1682 ATTGATGGTCTTCGAGGTAGTCTTAACCTGAAGTGTCTTCTCAGTTGTTCAGATAATGTT 1741  
|||||  
Db 60 ATTGATGGTCTTCGAGGTAGTCTTAACCTGAAGTGTCTTCTCAGTTGTTCAGATAATGTT 1

RESULT 8  
LOCUS SCU96099 1115 bp mRNA INV 15-APR-1999  
DEFINITION Sarcophaga crassipalpis 23kda heat shock protein SCHSP23 mRNA,  
complete cds.  
ACCESSION U96099  
VERSION U96099.2 GI:4584801  
KEYWORDS  
SOURCE Sarcophaga crassipalpis.  
ORGANISM Sarcophaga crassipalpis.  
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Oestroidea; Sarcophagidae; Sarcophaga.  
1 (bases 1 to 823)  
AUTHORS Flanagan,R.D., Tammariello,S.P., Joplin,K.H., Cikra-Ireland,R.A.,  
Yocum,G.D. and Denlinger,D.L.  
TITLE Diapause-specific gene expression in pupae of the flesh fly  
Sarcophaga crassipalpis  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (10), 5616-5620 (1998)  
MEDLINE 98245129  
REFERENCE 2 (bases 1 to 823)  
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.  
TITLE Upregulation of a 23 kda small heat shock protein transcript during  
pupal diapause in the flesh fly, Sarcophaga, crassipalpis  
JOURNAL Insect Biochem. Mol. Biol. 28 (9), 677-682 (1998)  
MEDLINE 98428060  
REFERENCE 3 (bases 1 to 823)  
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.  
TITLE Direct Submission  
JOURNAL Submitted (01-APR-1997) Entomology, Ohio State University, 1735  
MEDLINE Neil Ave, Columbus, OH 43210-1220, USA  
REFERENCE 4 (bases 1 to 1115)  
AUTHORS Yocum,G.D., Joplin,K.H. and Denlinger,D.L.

Direct Submission  
Submitted (15-APR-1999) Entomology, Ohio State University, 1735  
Neil Ave, Columbus, OH 43210-1220, USA  
REMARK Sequence, update by submitter  
COMMENT On Apr 15, 1999 this sequence version replaced gi:2058736.  
FEATURES Location/Qualifiers  
source 1. .1115  
/organism="Sarcophaga crassipalpis"  
/db\_xref="taxon:59312"  
88. .705  
/codon\_start=1  
/product="23kda heat shock protein SCHSP23"  
/protein\_id="AAC63387.1"  
/db\_xref="GI:2058737"  
/translation="MAGLLPLLSLIDELRESNFSYYGNDGFLGLSPYLTHROPQREP  
AHNLVGYSLPLSLSLRLNEHOVARRGEGKRGVSPVKGDFGFOVMDVAQFKFSELNV  
KVVDNCIVIEGKHEEREDOHGFVHFVRVLPKGYDADKVVSTLSSDGLVAVSVFK  
PQEEKSNERVIIQQTCPAHLNVKONTQEKVGANGEPQEEK"  
BASE COUNT 430 a 185 c 192 g 308 t  
ORIGIN

Query Match 1.7%; Score 47; DB 35; Length 1115;  
Best Local Similarity 100.0%; Pred. No. 7.2e-15;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2643 ACCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689  
|||||  
Db 1027 ACCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1073  
|||||

RESULT 9  
LOCUS RAT1TNAT 2305 bp mRNA ROD 27-APR-1993  
DEFINITION Rat L-type neutral amino acid transporter mRNA, complete cds.  
ACCESSION M77345  
VERSION M77345.1 GI:205238  
KEYWORDS L-type neutral amino acid transporter.  
SOURCE Rattus norvegicus (strain Wistar) cDNA to mRNA.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
REFERENCE 1 (bases 1 to 2305)  
AUTHORS Tate,S.S., Yan,N. and Udenfriend,S.  
TITLE Expression cloning of a Na+-independent neutral amino acid  
transporter from rat kidney  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89, 1-5 (1992)  
MEDLINE 92107996  
FEATURES Location/Qualifiers  
source 1. .2305  
/organism="Rattus norvegicus"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
39. .2090  
/codon\_start=1  
/product="L-type neutral amino acid transporter"  
/protein\_id="RAA41544.1"  
/translation="MNEKDKRDSIQMSMKGCTTNGFVNQEDIQEODPDSRDTPOSN  
AVSIDAPEPQLKVVRYAGMPKEVLFQSGQARYRPREILFWLVVSVFLLIGAT  
AIIISPKCLDMWQAGPMQIYPRSKDSDGNDGLKIQEKLDTALNKTIT  
WFSKPLDRFYAVEDEKIDPIFGTKMDFENLVAHVHDKLKIIDFIPNHTSKHP  
SQVSGTSRGKYTDYIHNCTHANGVTTPNNMLSVYGNSSQWDFEERKOCYFROL  
KEQPLNFRNPAVEIKEIKFLWSKGVDFSDAVKFLLEAKDLRNEIQVNTSQIP  
DIVTRYSEDIHFTTTQVGMHDLVRDFTQMNQFSREPRFPMGTVEVSAESTRIY  
YYGLSFIDEADFFNKYLATLDTLSGHTVYEAITSMMENPEGKPMNMWIGGSETSL  
TSRVGSYYNANMLLFTLPCTIYVYGEIGMGDISITINERYDITNALLSKSPMW  
DNSNAGTEAHTWLPNTSDYHTVNDVQKTPQSSALRYQDLQLSLHLLSGMW  
FCLLRDNNHVVYTRLEDIGDKVFLVNFGEESTVLNQETISDPTKRLRLSTNP  
C  
ASKGSDVDTHAVSLEKGEGLIIEHSMKTLHHQKAFRDKCFISNACYSVSSVLDLLYS  
C"

BASE COUNT 691 a 540 c 532 g 542 t  
ORIGIN

```

Query Match          1.7%:  Score 47:  DB 12:  Length 2305;
Best Local Similarity 100.0%:  Pred. No. 7.2e-15;
Matches 47:  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Oy 2643 ACCCCCAAAA*AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Db 2254 ACCCCCAAAA*AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2300

RESULT 10
AC013452 184553 bp DNA HTG 12-NOV-1999
LOCUS Homo sapiens chromosome 15 clone RP11-325E5 map 15q21.1, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
AC013452
AC013452.1 GI:6382485
KEYWORDS HTG; HTG5_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184553)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Sequencing of human chromosome 15 D155146-D155117region
Unpublished
2 (bases 1 to 184553)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
Direct Submission
Submitted (12-NOV-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
----- Genome Center
Center code: UWM5C
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@u.washington.edu
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Big Dye terminators
Assembly program: Phrap; version 0.990399.
* NOTE: This record contains 192 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 914: contig of 914 bp in length
1 915: gap of unknown length
1 1802: contig of 888 bp in length
1 1803: gap of unknown length
1 2360: contig of 558 bp in length
1 2361: gap of unknown length
1 3248: contig of 888 bp in length
1 3249: gap of unknown length
1 3769: contig of 521 bp in length
1 3770: gap of unknown length
1 4671: contig of 902 bp in length
1 4672: gap of unknown length
1 5662: contig of 991 bp in length
1 5663: contig of 899 bp in length

```

```

* 6562 7466: contig of 905 bp in length
* 7467 8367: contig of 901 bp in length
* 8368 9293: contig of 926 bp in length
* 9294 10195: contig of 902 bp in length
* 10196 10662: contig of 467 bp in length
* 10663 11812: contig of 1150 bp in length
* 11813 12734: contig of 922 bp in length
* 12735 13630: contig of 896 bp in length
* 13631 14561: contig of 931 bp in length
* 14562 15459: contig of 898 bp in length
* 15460 16461: contig of 1002 bp in length
* 16462 17358: contig of 897 bp in length
* 17359 18322: contig of 964 bp in length
* 18323 19222: contig of 900 bp in length
* 19223 20193: contig of 971 bp in length
* 20194 21221: contig of 1028 bp in length
* 21222 21727: contig of 506 bp in length
* 21728 22830: contig of 1103 bp in length
* 22831 23796: contig of 986 bp in length
* 23797 24970: contig of 1174 bp in length
* 24971 25769: contig of 799 bp in length
* 25770 26743: contig of 974 bp in length
* 26744 27692: contig of 949 bp in length
* 27693 28682: contig of 990 bp in length
* 28683 29675: contig of 993 bp in length
* 29676 30575: contig of 900 bp in length
* 30576 31555: contig of 980 bp in length
* 31556 32453: contig of 898 bp in length
* 32454 33425: contig of 972 bp in length
* 33426 34346: contig of 921 bp in length
* 34347 35510: contig of 1164 bp in length
* 35511 36420: contig of 910 bp in length
* 36421 36826: contig of 406 bp in length
* 36827 38207: contig of 1381 bp in length
* 38208 39027: contig of 820 bp in length
* 39028 39931: contig of 904 bp in length
* gap of unknown length

```

\* 39932 40863: contig of 932 bp in length  
\* gap of unknown length  
\* 40864 41894: contig of 1031 bp in length  
\* gap of unknown length  
\* 41895 42767: contig of 873 bp in length  
\* gap of unknown length  
\* 42768 43667: contig of 900 bp in length  
\* gap of unknown length  
\* 43668 44870: contig of 1203 bp in length  
\* gap of unknown length  
\* 44871 45843: contig of 973 bp in length  
\* gap of unknown length  
\* 45844 46783: contig of 940 bp in length  
\* gap of unknown length  
\* 46784 47851: contig of 1068 bp in length  
\* gap of unknown length  
\* 47852 48830: contig of 979 bp in length  
\* gap of unknown length  
\* 48831 50308: contig of 1478 bp in length  
\* gap of unknown length  
\* 50309 51547: contig of 1239 bp in length  
\* gap of unknown length  
\* 51548 52617: contig of 1070 bp in length  
\* gap of unknown length  
\* 52618 53551: contig of 934 bp in length  
\* gap of unknown length  
\* 53552 54449: contig of 898 bp in length  
\* gap of unknown length  
\* 54450 55434: contig of 985 bp in length  
\* gap of unknown length  
\* 55435 56625: contig of 1191 bp in length  
\* gap of unknown length  
\* 56626 57543: contig of 918 bp in length  
\* gap of unknown length  
\* 57544 58518: contig of 975 bp in length  
\* gap of unknown length  
\* 58519 59476: contig of 958 bp in length  
\* gap of unknown length  
\* 59477 60932: contig of 1456 bp in length  
\* gap of unknown length  
\* 60933 61859: contig of 927 bp in length  
\* gap of unknown length  
\* 61860 62750: contig of 891 bp in length  
\* gap of unknown length  
\* 62751 63683: contig of 933 bp in length  
\* gap of unknown length  
\* 63684 64581: contig of 898 bp in length  
\* gap of unknown length  
\* 64582 65562: contig of 981 bp in length  
\* gap of unknown length  
\* 65563 66461: contig of 899 bp in length  
\* gap of unknown length  
\* 66462 67594: contig of 1133 bp in length  
\* gap of unknown length  
\* 67595 69414: contig of 1820 bp in length  
\* gap of unknown length  
\* 69415 70498: contig of 1084 bp in length  
\* gap of unknown length  
\* 70499 71852: contig of 1354 bp in length  
\* gap of unknown length  
\* 71853 72767: contig of 915 bp in length  
\* gap of unknown length  
\* 72768 73669: contig of 902 bp in length  
\* gap of unknown length  
\* 73670 74584: contig of 915 bp in length  
\* gap of unknown length  
\* 74585 75486: contig of 902 bp in length  
\* gap of unknown length

Query Match 1.7%; Score 47; DB 43; Length 184553;  
Best Local Similarity 100.0%; Pred. No. 7e-15;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2643 ACCCCCCAAA 2689

Db 115012 ACCCCCCAAA 115058  
RESULT 11  
A69673  
LOCUS A69673 1305 bp DNA PAT 07-MAY-1999  
DEFINITION Sequence 3 from Patent WO9807872.  
ACCESSION A69673  
VERSION A69673.1 GI:4774286  
KEYWORDS Trichomonas vaginalis.  
SOURCE Trichomonas vaginalis.  
ORGANISM Eukaryota; Parabasalidea; Trichomonadida; Trichomonas.  
REFERENCE 1 (bases 1 to 1305)  
AUTHORS Campbell,R.S., Coombs,G.H., Mottram,J.C. and Pritchard,D.J.  
TITLE HOMOCYSTEINE DESULPHURASE FROM THE PROTOZOAN TRICHOMONAS VAGINALIS  
JOURNAL Patent: WO 9807872-A 26-FEB-1998;  
CAMPBELL ROBERT STEWART (GB)  
FEATURES  
source  
1..1305  
/organism="Trichomonas vaginalis"  
/db\_xref="taxon:5722"  
/cell\_lines="G3"  
/clone="MGL2"  
16..1212  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAB42512.1"  
/db\_xref="GI:4774287"  
/translation="MSGHAIPTHTDLSIHANPKDQFGAIVAPIYQTSTELFNCD  
OGGAREFGKEAGYMYTRIGNPTNSALEGKIAKLEHAEACAAATASGMGAIAASVWTFLK  
AGDHLISDDCLYGCTHALFEHOLRKFGVEVDFDMAVPGNIEKHLKPNTRIVYFETPA  
NPTLVIEDIAVKQARKQKQDILVIVONTFASPLNPLDLCVDIVVHSATKYNGHT  
DVVAGVCSRADIIAKVKSQIKDITGAILISPDADWLITRGILITLDMRVRAAEARQK  
VAEFLHEHKAVKVVYPGLPDHPGHEIAKKMRMFGSMIAFDVDGLEGAKKVVLDNCHV  
VSLAYSLGGPESLIQHPASMTAGVPKPEEREAAGLTDNLRLSVGCENYQDIIDDLKQ  
ALLDLI"

BASE COUNT 374 a 376 c 269 g 286 t  
ORIGIN

Query Match 1.7%; Score 45; DB 5; Length 1305;  
Best Local Similarity 100.0%; Pred. No. 2.6e-14;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2644 CCCCCAAA 2689  
Db 1257 CCCCCAAA 1302

RESULT 12  
A69681  
LOCUS A69681 1306 bp DNA PAT 07-MAY-1999  
DEFINITION Sequence 11 from Patent WO9807872.  
ACCESSION A69681  
VERSION A69681.1 GI:4774294  
KEYWORDS unidentified.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1306)  
AUTHORS Campbell,R.S., Coombs,G.H., Mottram,J.C. and Pritchard,D.J.  
TITLE HOMOCYSTEINE DESULPHURASE FROM THE PROTOZOAN TRICHOMONAS VAGINALIS  
JOURNAL Patent: WO 9807872-A 26-FEB-1998;  
CAMPBELL ROBERT STEWART (GB)  
FEATURES  
source  
1..1306  
/organism="unidentified"  
/db\_xref="taxon:32644"  
17..1213  
/note="unnamed protein product"  
/codon\_start=1

```
/protein_id="CAB42514.1"
/db_xref="GI:4774295"
/translacion="MSGHALDPTHDTLSIHANPKQDQFGAIVAPITYOTSTFLDNCD
QGGRFGKAGWYTRIGNPTNSALEGKIAKLEHAECATAGSMGAIAASVMTFLK
AGHLISDDGLYCTHALFEHQURKEGVEVDFIDMAVPGNIEKHLKPNTRIVYETPA
NPTLKVIDIDAYKQARKDILVIYDNTFASPILNPLDGLGVDIVHRSATKYINGHT
DVAGLVKCAKVDIAKVKQKIDITGAIISPHDAWLITRGTLTIDMRKVAENAOAK
VAEPLHEHKAVKVVYVGLPDHPGHEIAKKOMKFGSMIAEDVDGLEKAKKVLONCHV
VSLAVSLGGPESLIQHPASMTWAGVPKEEREAAGLTDLNLRSLVGCENVQDIIDDLAK
ALDLVL"
BASE COUNT      374 a   376 c   271 g   285 t
ORIGIN

Query Match      1.7%; Score 46; DB 5; Length 1306;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||||.....
DB 1258 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1303
      |||||.....

RESULT 13
AF067420      1651 bp      mRNA      PRI      10-JUN-1998
LOCUS      Homo sapiens SNC73 protein (SNC73) mRNA, complete cds.
DEFINITION
ACCESSION      AF067420
VERSION      AF067420.1 GI:3201899
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1651)
Zhang S., Cao J., Cao W., Cai, X. and Geng, L.
Identification and characterization of SNC73, a gene which is
down-regulated in colorectal cancer
Unpublished
2 (bases 1 to 1651)
Zhang S., Cao J., Cao W., Cai, X. and Geng, L.
Direct Submission
Submitted (19-MAY-1998) Cancer Institute, Zhejiang Medical
University, 68 Jiefang Road, Hangzhou, Zhejiang 310009, P. R. China
3 (bases 1 to 1651)
Zhang S., Cao J., Cao W., Cai, X. and Geng, L.
Direct Submission
Submitted (09-JUN-1998) Cancer Institute, Zhejiang Medical
University, 68 Jiefang Road, Hangzhou, Zhejiang 310009, P. R. China
Sequence update by submitter
REMARK      On Jun 10, 1998 this sequence version replaced gi:3192828.
FEATURES
source
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /note="sequence obtained by subtractive hybridization
    between normal mucosa and colorectal cancer tissue"
    1..1651
    /gene="SNC73"
    396..1550
    /gene="SNC73"
    /codon_start=1
    /product="SNC73 protein"
    /protein_id="AAC19365.1"
    /db_xref="GI:3201900"
    /translacion="MAVYICVSGGIYDSSGPDYWGQGLTVTSASPTSPKVPPLSL
    CSTQPDNVIAGLVFPQPLSVNTSSEGGVTARFPDQDAGDLYTTSQLT
    LPTQCLAGKSVTCHVKHTNPSQDVTPCPVPSTPTPTPTPTPTPTPTPTPTPTPT
    LRPALEDLLGSAKNTCTLTGLRASGVTFWTPTSSGKSAVQGPPTDRLCGCVSVSS
    VLPCCAPNHHGKTFCTAAYPESKTPLTATLSKGNTRPEVLHPPPSLALNEL
    VTLTCLARGE SPKDLVLRVLOGSQELPRLTASROEPGQTTTAVTSLIRVAE
    DWKGDGDFSCMVGHEALPLAFTQKTDRLAGRTHVNVSVNAEVDGTCY"
BASE COUNT      376 a   516 c   443 g   316 t

Query Match      1.7%; Score 46; DB 11; Length 1651;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||||.....
DB 1590 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1635
      |||||.....

RESULT 14
HSU75370      3832 bp      mRNA      PRI      23-MAY-1997
LOCUS      Human mitochondrial RNA polymerase mRNA, nuclear gene encoding
DEFINITION      mitochondrial protein, complete cds.
ACCESSION      U75370
VERSION      U75370.1 GI:2114395
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3832)
Tiranti, V., Savola, A., Forti, F., D'Apolito, M.F., Centra, M.,
Rocchi, M. and Zeviani, M.
Identification of the gene encoding the human mitochondrial RNA
polymerase (h-mtRPQD) by cyberscreening of the Expressed Sequence
Tags database
Hum. Mol. Genet. 6 (4), 615-625 (1997)
2 (bases 1 to 3832)
Tiranti, V., D'Apolito, M.F., Forti, F., Rocchi, M., Savola, A. and
Zeviani, M.
Direct Submission
Submitted (18-OCT-1996) Molecular Medicine, Children's Hospital
Bambino Gesù - IRCCS, Piazza S. Onofrio 4, Rome, RM 00165, Italy
FEATURES
source
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /tissue_type="liver"
    /chromosome="19"
    /map="19p13.3"
    33..3725
    /codon_start=1
    /product="mitochondrial RNA polymerase"
    /protein_id="AAB58255.1"
    /db_xref="GI:2114396"
    /translacion="MSALCWGGAAGLKRALRPGCPGLPGKEGTAGGVCGRSSSA
    SPQEQDQDRRKDGHVLELVQARVQLQASVSEVVNRVDVARLPCGSGDGLQ
    PRKVMGAKDATPVCGRWAKILEKDKRTQQMRMORLAKLQMPQSGEFALTRRL
    QVEPRLLSKOMAGCLECTROAPESPQELARLLOEAPGKLSLDVEQAPSGHQAQ
    LSGQORLLAFFKCLLTDLPLAHLILVYHHQQRKRLLLIDHTNAYLWAGQA
    FKELVYLVFKDAGLTPDLLSTAALQCMGQDQDAGTIERCLDMSQSEGLKALF
    TAVLSEDRATYLVKAVHKVPTFSLPPLPPVNTSKLRLDYARDQVRSYKULPL
    LKTLQCFEKLHMLASRYCVSVSEKPTLPSKEVHKARKTLKTRDQWREKALRLR
    FSRHVQRQVSGVQALQNHRYKYLCLLASDAEVPQCLPQYWEELGAPALREQP
    WPLPVOMELGKLLAEMLQATOMPQSLDKPRHSRLVPVLYHYVSEFNRVQOIGILKPH
    PAYVOLLKAEPLTTEAVDVPMLQPLPPTSPHSQAFLLSPTLMRTVEGATQOE
    LLETCPPTALHGALDALTQNCANWNGRVLVLVQLQAGCCGCGPQVAPPSAPQ
    PPEAHLPHSAAPARKAEKRLRELAKCKYKREMSRLKALRLSLAQHLRLRVFWLPH
    NMDRFRTYPCPFHNLGSDVAFALLEAQGRPLGPHGDWLKILHVLNLTGKKRFP
    LRKLAAEEYDDILDADQPLTGRRWMGAEPEWOTLACCNEVANAVRASDPAAY
    SHLPVHDGSGCGLQHYAALGRDSGAASVNLSPDQDYSVGAAYQVEFRRDQAY
    RGMVAQVLESFTRKVKQTVMTVYGVTRYSGRIQIEKRLSEDSPEQEFWEDASH
    YLVQVFKSLQEMESGTRAIOWHLTSEARLISHMSVWEVMTPLGVPIQVYRLDSKY
    KOIGQSGTITHTNGDISRKPNTRKQKGFPPNFTHSLDSSHMLTALHCYKGLTF
    VSHDVCVTHAADVYVMNOVCBQFVRLHSEPLQLDSRLFKRCSEPEQKILEASQL
    KETIQVKKPGAEDEQVKSTYFS"
BASE COUNT      725 a   1235 c   1259 g   613 t
```

```

ORIGIN
Query Match      1.7%; Score 46; DB 10; Length 3832;
Best Local Similarity 100.0%; Pred. No. 2.6e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Ddb 3771 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3816
|||||

RESULT 15
AF088979 12971 bp DNA INV 01-SEP-1999
Dictyostelium discoideum beige protein homolog (lvsA) gene.
complete cds.
AF088979
AF088979.1 GI:5814090
Dictyostelium discoideum.
Dictyostelium discoideum.
Eukaryota; Dictyostelidia; Dictyostelium.
1 (bases 1 to 12971)
Kwak,E., Gerald,N., Laroche,H., D.A., Vithalani,K.K., Niswonger,M.
and De Lozanne,A.
lvsA, a protein related to the mouse beige protein, is essential
for cytokinesis in Dictyostelium
Unpublished
2 (bases 1 to 12971)
De Lozanne,A.
Direct Submission
Submitted (31-Aug-1998) Cell Biology, Duke University Medical
Center, 367 Nanaline Duke Building, Durham, NC 27710, USA
Location/Qualifiers
1..12971
/gene=lvsA
/organism=Dictyostelium discoideum
/strain=Dhl
/db_xref=taxon:44689
join(1..2867,3029..3217,3345..3597,3724..3903,3959..12971)
/gene=lvsA
/product=beige protein homolog
1..1451
/gene=lvsA
1..12971
/gene=lvsA
join(1452..2867,3029..3217,3345..3597,3724..3903,
3959..12603)
/gene=lvsA
/note=LVSa; affected in the REMI cytokinesis mutant
60tc5
/codon_start=1
/product=beige protein homolog
/protein_id=AAD52096.1
/db_xref=GI:5814091
/translation=MFRFKDLIGTSDPTQPVPHSPGHPHPPQQOQQOQQOQQO
QOOOQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
KSKDKVMRLNKLPLVEGLYEDKIDTKPMVDIFGNNSKSFSAFLRVLKDINEI
MKANSPOOPPOPQLIKETIAELVFESTTSQSGSFGFLLYSIELSESNTSCAE
MAEASIPMLVRKYDFLVPYTMMETTGILEEKLRITLCFLSKQSAIELOKTD
TLSTLFALMSNECPHSRRPVAKRVGSLELDLPITYIIINSKRVIINIKDLYNY
FYMTSESVILCRITIKILSSSKSTILLDEFQRNDGYFLVDSFRLESSKKPAL
FEQLDSIVLVGYGVNPNPNAKFLERYFLKSNEYENRVKLDRLLSYSSNT
NPSFTLEMINKLTIOEFYESLGKHYHVKIVCFVTNLCPVFQELSTSLVGE
NPFVTEMLNLTIVNFPRKHIFRETGLDILVKVIDIAODIRLNKNKGSG
SNGPIVCMCTGKEADSNDQALSIIKVESFOILLDSLFIESPONISILRNIN
IFIETISSFINILLRELPSVGKSLRILOLIKVDPEPTOKEFGLIKVLTVSNK
MEKLELIESICRCITTSALCQMNIINRENFEOQIGYKTFFESCFLVESFSFNKDRNRWD
INFPMKSIDNALRKLFNISKAHDSPREHGFWVLSVFISSLESFSFNKDRNRWD
IKFDVMTENLASDOIQNOMIINNVSFNVLIDIPIHENKDFRLQJHISRNKWAEGYR
YNQALSKLIPDWILSRFPNSLNANDLPQLLSLIQTGVGANCISGSLRFQVKLL
OPENSHVELLKILSMASKSPPTPYFFENISKIPFGYIRVIPITERAWPPTYIMFW

```

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 28862)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 5  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 28862)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-SEP-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 www.jgi.doe.gov.  
 COMMENT \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 32 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 725: contig of 725 bp in length  
 \* gap of unknown length  
 \* 726 1942: contig of 1217 bp in length  
 \* gap of unknown length  
 \* 1943 2512: contig of 570 bp in length  
 \* gap of unknown length  
 \* 2513 5735: contig of 3223 bp in length  
 \* gap of unknown length  
 \* 5736 6314: contig of 779 bp in length  
 \* gap of unknown length  
 \* 6315 7081: contig of 567 bp in length  
 \* gap of unknown length  
 \* 7082 7787: contig of 706 bp in length  
 \* gap of unknown length  
 \* 7788 8551: contig of 764 bp in length  
 \* gap of unknown length  
 \* 8552 9198: contig of 647 bp in length  
 \* gap of unknown length  
 \* 9199 9983: contig of 785 bp in length  
 \* gap of unknown length  
 \* 9984 10428: contig of 445 bp in length  
 \* gap of unknown length  
 \* 10429 10686: contig of 258 bp in length  
 \* gap of unknown length  
 \* 10687 11348: contig of 862 bp in length  
 \* gap of unknown length  
 \* 11349 12265: contig of 717 bp in length  
 \* gap of unknown length  
 \* 12266 13168: contig of 903 bp in length  
 \* gap of unknown length  
 \* 13169 13854: contig of 686 bp in length  
 \* gap of unknown length  
 \* 13855 14636: contig of 782 bp in length  
 \* gap of unknown length  
 \* 14637 15872: contig of 1236 bp in length  
 \* gap of unknown length  
 \* 15873 16842: contig of 970 bp in length  
 \* gap of unknown length  
 \* 16843 17579: contig of 737 bp in length  
 \* gap of unknown length  
 \* 17580 18803: contig of 1224 bp in length  
 \* gap of unknown length  
 \* 18804 19386: contig of 583 bp in length  
 \* gap of unknown length  
 \* 19387 20433: contig of 1047 bp in length  
 \* gap of unknown length  
 \* 20434 21170: contig of 737 bp in length  
 \* gap of unknown length  
 \* 21171 21263: contig of 93 bp in length  
 \* gap of unknown length  
 \* 21264 21913: contig of 630 bp in length

\* 21914 23120: contig of 1207 bp in length  
 \* gap of unknown length  
 \* 23121 24749: contig of 1629 bp in length  
 \* gap of unknown length  
 \* 24750 25568: contig of 819 bp in length  
 \* gap of unknown length  
 \* 25569 26585: contig of 1017 bp in length  
 \* gap of unknown length  
 \* 26586 27627: contig of 1042 bp in length  
 \* gap of unknown length  
 \* 27628 28862: contig of 1235 bp in length.  
 FEATURES Location/Qualifiers  
 source 1..28862  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CIT-HSPC\_349C3"  
 BASE COUNT 7289 a 7328 c 7041 g 6799 t 405 others  
 ORIGIN  
 Query Match 1.7%; Score 46; DB 41; Length 28862;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2644 CCCCCAAA 2689  
 |||||  
 Db 1420 CCCCCAAA 1465  
 |||||  
 RESULT 17  
 AC016250/78250 bp DNA HTG 23-NOV-1999  
 LOCUS Homo sapiens chromosome 15 clone RP11-516P9 map 15, LOW-PASS  
 DEFINITION SEQUENCE SAMPLING.  
 AC016250  
 VERSION AC016250.1 GI:6466544  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 78250)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 78250)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Gage,D., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boquelavkiy,L., Boukhalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C.,  
 Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehotzky,J.,  
 Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,  
 McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V.,  
 Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,  
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A.,  
 Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/BM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submission@genome.wi.mit.edu



```
----- Project Information
Center project name: L4978
Center clone name: 516_P_9
-----
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 981: contig of 981 bp in length
*      gap of unknown length
*      982 1969: contig of 988 bp in length
*      gap of unknown length
*      1970 2951: contig of 982 bp in length
*      gap of unknown length
*      2952 3967: contig of 1016 bp in length
*      gap of unknown length
*      3968 4876: contig of 909 bp in length
*      gap of unknown length
*      4877 5699: contig of 823 bp in length
*      gap of unknown length
*      5700 6669: contig of 970 bp in length
*      gap of unknown length
*      6670 7627: contig of 958 bp in length
*      gap of unknown length
*      7628 8552: contig of 925 bp in length
*      gap of unknown length
*      8553 9533: contig of 981 bp in length
*      gap of unknown length
*      9534 10496: contig of 963 bp in length
*      gap of unknown length
*      10497 11483: contig of 987 bp in length
*      gap of unknown length
*      11484 12457: contig of 974 bp in length
*      gap of unknown length
*      12458 13440: contig of 983 bp in length
*      gap of unknown length
*      13441 14419: contig of 979 bp in length
*      gap of unknown length
*      14420 15392: contig of 973 bp in length
*      gap of unknown length
*      15393 16338: contig of 946 bp in length
*      gap of unknown length
*      16339 17295: contig of 957 bp in length
*      gap of unknown length
*      17296 18288: contig of 993 bp in length
*      gap of unknown length
*      18289 19227: contig of 939 bp in length
*      gap of unknown length
*      19228 20137: contig of 910 bp in length
*      gap of unknown length
*      20138 21111: contig of 974 bp in length
*      gap of unknown length
*      21112 22088: contig of 977 bp in length
*      gap of unknown length
*      22089 23089: contig of 1001 bp in length
*      gap of unknown length
*      23090 24039: contig of 950 bp in length
*      gap of unknown length
*      24040 24951: contig of 912 bp in length
*      gap of unknown length
*      24952 25924: contig of 973 bp in length
*      gap of unknown length
*      25925 26930: contig of 1006 bp in length
*      gap of unknown length
*      26931 27913: contig of 983 bp in length
*      gap of unknown length
```

```
27914 28882: contig of 969 bp in length
*      gap of unknown length
*      28883 29825: contig of 943 bp in length
*      gap of unknown length
*      29826 30825: contig of 1000 bp in length
*      gap of unknown length
*      30826 31777: contig of 952 bp in length
*      gap of unknown length
*      31778 32754: contig of 977 bp in length
*      gap of unknown length
*      32755 33733: contig of 979 bp in length
*      gap of unknown length
*      33734 34725: contig of 992 bp in length
*      gap of unknown length
*      34726 35667: contig of 942 bp in length
*      gap of unknown length
*      35668 36638: contig of 971 bp in length
*      gap of unknown length
*      36639 37590: contig of 952 bp in length
*      gap of unknown length
*      37591 38548: contig of 958 bp in length
*      gap of unknown length
*      38549 39502: contig of 954 bp in length
*      gap of unknown length
*      39503 40448: contig of 946 bp in length
*      gap of unknown length
*      40449 41431: contig of 983 bp in length
*      gap of unknown length
*      41432 42427: contig of 996 bp in length
*      gap of unknown length
*      42428 43424: contig of 997 bp in length
*      gap of unknown length
*      43425 44355: contig of 931 bp in length
*      gap of unknown length
*      44356 45315: contig of 980 bp in length
*      gap of unknown length
*      45316 46264: contig of 949 bp in length
*      gap of unknown length
*      46265 47222: contig of 958 bp in length
*      gap of unknown length
*      47223 48185: contig of 963 bp in length
*      gap of unknown length
*      48186 49143: contig of 958 bp in length
*      gap of unknown length
*      49144 50119: contig of 976 bp in length
*      gap of unknown length
*      50120 51091: contig of 972 bp in length
*      gap of unknown length
*      51092 52032: contig of 941 bp in length
*      gap of unknown length
*      52033 52985: contig of 953 bp in length
*      gap of unknown length
*      52986 53943: contig of 958 bp in length
*      gap of unknown length
*      53944 54947: contig of 1004 bp in length
*      gap of unknown length
*      54948 55933: contig of 986 bp in length
*      gap of unknown length
*      55934 56879: contig of 946 bp in length
*      gap of unknown length
*      56880 57886: contig of 1007 bp in length
*      gap of unknown length
*      57887 58879: contig of 993 bp in length
*      gap of unknown length
*      58880 59864: contig of 985 bp in length
*      gap of unknown length
*      59865 60834: contig of 970 bp in length
*      gap of unknown length
*      60835 61817: contig of 983 bp in length
*      gap of unknown length
*      61818 62801: contig of 984 bp in length
*      gap of unknown length
*      62802 63787: contig of 986 bp in length
```





\* 43854 44494: contig of 641 bp in length  
 \* gap of unknown length  
 \* 44495 45122: contig of 628 bp in length  
 \* gap of unknown length  
 \* 45123 45775: contig of 653 bp in length  
 \* gap of unknown length  
 \* 45776 46407: contig of 632 bp in length  
 \* gap of unknown length  
 \* 46408 47096: contig of 689 bp in length  
 \* gap of unknown length  
 \* 47097 47746: contig of 650 bp in length  
 \* gap of unknown length  
 \* 47747 48434: contig of 688 bp in length  
 \* gap of unknown length  
 \* 48435 49043: contig of 609 bp in length  
 \* gap of unknown length  
 \* 49044 49487: contig of 444 bp in length  
 \* gap of unknown length  
 \* 49488 50164: contig of 677 bp in length  
 \* gap of unknown length  
 \* 50165 51352: contig of 1188 bp in length  
 \* gap of unknown length  
 \* 51353 52008: contig of 656 bp in length  
 \* gap of unknown length  
 \* 52009 53126: contig of 1118 bp in length  
 \* gap of unknown length  
 \* 53127 53880: contig of 754 bp in length  
 \* gap of unknown length  
 \* 53881 54583: contig of 703 bp in length  
 \* gap of unknown length  
 \* 54584 55426: contig of 843 bp in length  
 \* gap of unknown length  
 \* 55427 55899: contig of 473 bp in length  
 \* gap of unknown length  
 \* 55900 56462: contig of 563 bp in length  
 \* gap of unknown length  
 \* 56463 57089: contig of 627 bp in length  
 \* gap of unknown length  
 \* 57090 57665: contig of 576 bp in length  
 \* gap of unknown length  
 \* 57666 58483: contig of 818 bp in length  
 \* gap of unknown length  
 \* 58484 59706: contig of 1223 bp in length  
 \* gap of unknown length  
 \* 59707 60945: contig of 1239 bp in length  
 \* gap of unknown length  
 \* 60946 61533: contig of 588 bp in length  
 \* gap of unknown length  
 \* 61534 62096: contig of 563 bp in length  
 \* gap of unknown length  
 \* 62097 62774: contig of 678 bp in length  
 \* gap of unknown length  
 \* 62775 63387: contig of 613 bp in length  
 \* gap of unknown length  
 \* 63388 63834: contig of 447 bp in length  
 \* gap of unknown length  
 \* 63835 64293: contig of 459 bp in length  
 \* gap of unknown length  
 \* 64294 64828: contig of 535 bp in length  
 \* gap of unknown length  
 \* 64829 65563: contig of 735 bp in length  
 \* gap of unknown length  
 \* 65564 65984: contig of 421 bp in length  
 \* gap of unknown length  
 \* 65985 66632: contig of 648 bp in length  
 \* gap of unknown length  
 \* 66633 67220: contig of 588 bp in length  
 \* gap of unknown length  
 \* 67221 67980: contig of 760 bp in length  
 \* gap of unknown length  
 \* 67981 68659: contig of 679 bp in length  
 \* gap of unknown length  
 \* 68660 69169: contig of 510 bp in length

\* 69170 69921: contig of 752 bp in length  
 \* gap of unknown length  
 \* 69922 70656: contig of 735 bp in length  
 \* gap of unknown length  
 \* 70657 71341: contig of 685 bp in length  
 \* gap of unknown length  
 \* 71342 72143: contig of 802 bp in length  
 \* gap of unknown length  
 \* 72144 72754: contig of 611 bp in length  
 \* gap of unknown length  
 \* 72755 73453: contig of 699 bp in length  
 \* gap of unknown length  
 \* 73454 74160: contig of 707 bp in length  
 \* gap of unknown length  
 \* 74161 74824: contig of 664 bp in length  
 \* gap of unknown length  
 \* 74825 75738: contig of 914 bp in length  
 \* gap of unknown length  
 \* 75739 76434: contig of 696 bp in length  
 \* gap of unknown length  
 \* 76435 76938: contig of 504 bp in length  
 \* gap of unknown length  
 \* 76939 77655: contig of 717 bp in length  
 \* gap of unknown length  
 \* 77656 78866: contig of 1211 bp in length  
 \* gap of unknown length  
 \* 78867 80003: contig of 1137 bp in length  
 \* gap of unknown length  
 \* 80004 80833: contig of 830 bp in length  
 \* gap of unknown length  
 \* 80834 81845: contig of 1012 bp in length  
 \* gap of unknown length  
 \* 81846 82882: contig of 1037 bp in length  
 \* gap of unknown length  
 \* 82883 83991: contig of 1109 bp in length  
 \* gap of unknown length  
 \* 83992 86061: contig of 2070 bp in length  
 \* gap of unknown length  
 \* 86062 86816: contig of 755 bp in length  
 \* gap of unknown length  
 \* 86817 88271: contig of 1455 bp in length  
 \* gap of unknown length  
 \* 88272 89289: contig of 1018 bp in length  
 \* gap of unknown length  
 \* 89290 90069: contig of 780 bp in length  
 \* gap of unknown length  
 \* 90070 91133: contig of 1064 bp in length  
 \* gap of unknown length  
 \* 91134 91833: contig of 700 bp in length  
 \* gap of unknown length  
 \* 91834 92661: contig of 828 bp in length  
 \* gap of unknown length  
 \* 92662 93414: contig of 753 bp in length  
 \* gap of unknown length  
 \* 93415 94471: contig of 1057 bp in length  
 \* gap of unknown length  
 \* 94472 95622: contig of 1151 bp in length  
 \* gap of unknown length  
 \* 95623 97279: contig of 1657 bp in length  
 \* gap of unknown length  
 \* 97280 98246: contig of 967 bp in length  
 \* gap of unknown length  
 \* 98247 99518: contig of 1272 bp in length  
 \* gap of unknown length  
 \* 99519 100837: contig of 1319 bp in length  
 \* gap of unknown length  
 \* 100838 102229: contig of 1392 bp in length  
 \* gap of unknown length  
 \* 102230 103965: contig of 1736 bp in length  
 \* gap of unknown length  
 \* 103966 105694: contig of 1729 bp in length  
 \* gap of unknown length

```
Query Match      1.7%; Score 46; DB 41; Length 108064;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 17281 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 17326

RESULT 20
AC017014.1/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC017014 Accession AC017014
Fragment Name      Begin      End
AC017014_0         1      110000
AC017014_1        100001    210000
AC017014_2        200001    310000
AC017014_3        300001    367220
Continuation (2 of 4) of AC017014 from base 100001 (AC017014 Homo sapiens clone RP11-142)

Query Match      1.7%; Score 46; DB 45; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 34117 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 34072

RESULT 21
AC009889/c
LOCUS
DEFINITION
Homo sapiens clone H_NH0113K19, *** SEQUENCING IN PROGRESS ***, 13
unordered pieces.
AC009889
VERSION
AC009889.1 GI:5822964
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132333)
McCombie,W.R.
Human Genomic Sequence
Unpublished
2 (bases 1 to 132333)
McCombie,W.R.
Direct Submission
Submitted (04-SEP-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 21380: contig of 21380 bp in length
* 21381
* 21685: gap of unknown length
* 21686
* 40255: contig of 18570 bp in length
* 40256
* 40561: gap of unknown length
* 40561
* 57759: contig of 17199 bp in length
* 57760
* 58064: gap of unknown length
* 58065
* 72684: contig of 14620 bp in length
* 72685
* 72989: gap of unknown length
* 72990
* 87082: contig of 14093 bp in length
* 87083
* 87387: gap of unknown length
* 87388
* 98798: contig of 11411 bp in length
* 98799
* 99102: gap of unknown length
```

```
* 99103 108556: contig of 9554 bp in length
* 10857 108960: gap of unknown length
* 108961 116080: contig of 7120 bp in length
* 116081 116384: gap of unknown length
* 116385 120735: contig of 4351 bp in length
* 120736 121039: gap of unknown length
* 121040 124805: contig of 3766 bp in length
* 124806 125110: gap of unknown length
* 125110 128610: contig of 3501 bp in length
* 128611 128914: gap of unknown length
* 128915 131764: contig of 2850 bp in length
* 131765 132069: gap of unknown length
* 132069 132333: contig of 265 bp in length.
FEATURES
Location/Qualifiers
1..132333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="H_NH0113K19"
BASE COUNT 37630 a 26434 c 26555 g 38040 t 3674 others
ORIGIN

Query Match      1.7%; Score 46; DB 41; Length 132333;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Db 131230 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 131185

RESULT 22
AL133259/c
LOCUS
DEFINITION
Homo sapiens chromosome 6 clone RP1-258E1, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
AL133259
ACCESSION
AL133259.1 GI:6522492
VERSION
HTG; HTGS_PHASE1.
KEYWORDS
human.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140513)
Sims,S.
Direct Submission
Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00004 Length: 38565p
Contig_ID: 00036 Length: 1729bp
Contig_ID: 00065 Length: 1361bp
Contig_ID: 00101 Length: 3279bp
Contig_ID: 00107 Length: 4463bp
Contig_ID: 00133 Length: 9012bp
Contig_ID: 00136 Length: 5087bp
Contig_ID: 00138 Length: 4932bp
Contig_ID: 00166 Length: 3138bp
Contig_ID: 00185 Length: 4561bp
Contig_ID: 00204 Length: 2642bp
Contig_ID: 00271 Length: 7384bp
Contig_ID: 00341 Length: 3289bp
Contig_ID: 00344 Length: 4762bp
Contig_ID: 00346 Length: 1650bp
Contig_ID: 00419 Length: 3620bp
Contig_ID: 00457 Length: 1422bp
Contig_ID: 00472 Length: 3681bp
```

Contig_ID:	00515	Length:	2744bp
Contig_ID:	00520	Length:	2352bp
Contig_ID:	00530	Length:	1203bp
Contig_ID:	00601	Length:	8631bp
Contig_ID:	00644	Length:	4896bp
Contig_ID:	00662	Length:	7143bp
Contig_ID:	00679	Length:	1754bp
Contig_ID:	00697	Length:	1162bp
Contig_ID:	00700	Length:	1211bp
Contig_ID:	00850	Length:	3517bp
Contig_ID:	00867	Length:	2919bp
Contig_ID:	00878	Length:	1163bp
Contig_ID:	00894	Length:	5149bp
Contig_ID:	00912	Length:	1801bp.

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

Location/Qualifiers

```

1..140513
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP1-258E1"
 /clone_lib="RPC1-1"

```

BASE COUNT	32737 a	24991 c	24342 g	33637 t	24806 others
ORIGIN					

Query Match            1.7%; Score 46; DB 33; Length 140513;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2644	CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
Dn	119775	CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 119730

---

RESULT	23
HSDJ315G1	
LOCUS	HSDJ315G1 144387 bp DNA HTG 27-NOV-1999
DEFINITION	Homo sapiens chromosome X clone RP1-315G1 map q24-q25, *** SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION	AL121601
VERSION	AL121601.3 Gi:6478115
KEYWORDS	HTG; HTGS-PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 144387) Pavitt,R. Direct Submission Submitted (27-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerquest@sanger.ac.uk On Nov 29, 1999 this sequence version replaced gi:6066979. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Contig_ID: 00024 Length: 1179bp Contig_ID: 00226 Length: 1287bp Contig_ID: 00291 Length: 7809bp Contig_ID: 00308 Length: 10670bp Contig_ID: 00334 Length: 23720bp Contig_ID: 00341 Length: 5230bp Contig_ID: 00491 Length: 6598bp Contig_ID: 00648 Length: 6405bp Contig_ID: 00949 Length: 26471bp Contig_ID: 01110 Length: 5629bp
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

Best Local Similarity 100.0%; Pred. No. 2.5e-14; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAA 2689  
 Db 668 CCCCCCAA 713

RESULT 25  
 AC011280  
 LOCUS Homo sapiens clone 13\_K\_8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 3  
 DEFINITION AC011280 163533 bp DNA HTG  
 AC011280  
 ACCESSION Homo sapiens  
 VERSION Homo sapiens clone 13\_K\_8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 3  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens, clone 13\_K\_8  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 163533)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Casale,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hags,B., Hearford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEvan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE Direct Submission  
 JOURNAL Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On Nov 5, 1999 this sequence version replaced 91:601210.  
 All repeats were identified using RepeatMasker: Smit, A.F.A. &  
 Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 20081: contig of 20081 bp in length  
 \* gap of unknown length  
 \* 20082 50254: contig of 30173 bp in length  
 \* gap of unknown length  
 \* 50255 163533: contig of 113279 bp in length.  
 \* Location/Qualifiers  
 1. 163533  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="13\_K\_8"  
 /clone\_lib="RPC1-11 Human Male BAC"

BASE COUNT 53421 a 30093 c 29840 g 50177 t 2 others  
 ORIGIN

Query Match 1.7%; Score 46; DB 43; Length 163533;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
 Mismatches 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAA 2689  
 Db 20431 CCCCCCAA 20476

RESULT 26  
 AC012067  
 LOCUS Homo sapiens clone NH0192N14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4  
 DEFINITION AC012067 165284 bp DNA HTG  
 AC012067  
 ACCESSION Homo sapiens  
 VERSION Homo sapiens clone NH0192N14, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 4  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165284)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 8528: contig of 8528 bp in length  
 \* 8529 8545: gap of unknown length  
 \* 8546 49205: contig of 40660 bp in length  
 \* 49206 49222: gap of unknown length  
 \* 49223 99112: contig of 49890 bp in length  
 \* 99113 99129: gap of unknown length  
 \* 99130 165284: contig of 66155 bp in length.  
 \* Location/Qualifiers  
 1. 165284  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NH0192N14"

BASE COUNT 53579 a 29550 c 29161 g 52943 t 51 others  
 ORIGIN

Query Match 1.7%; Score 46; DB 42; Length 165284;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
 Mismatches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAA 2689  
 Db 413 CCCCCCAA 458

RESULT 27  
 AC012457/c  
 LOCUS Homo sapiens clone NH0509E16, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 17  
 DEFINITION AC012457  
 ACCESSION Homo sapiens  
 VERSION Homo sapiens clone NH0509E16, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 17  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 165414)

AUTHORS Waterston, R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165414)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT  
 \* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2196: contig of 2196 bp in length  
 \* 2197  
 \* 2214: gap of unknown length  
 \* 2215 4460: contig of 2246 bp in length  
 \* 4461 4478: gap of unknown length  
 \* 4479 8425: contig of 3947 bp in length  
 \* 8426 8443: gap of unknown length  
 \* 8444 12603: contig of 4160 bp in length  
 \* 12604 12621: gap of unknown length  
 \* 12622 18021: contig of 5400 bp in length  
 \* 18022 18039: gap of unknown length  
 \* 18040 22844: contig of 4805 bp in length  
 \* 22845 22862: gap of unknown length  
 \* 22863 27028: contig of 4166 bp in length  
 \* 27029 27046: gap of unknown length  
 \* 27047 33961: contig of 6915 bp in length  
 \* 33962 33979: gap of unknown length  
 \* 33980 42011: contig of 8032 bp in length  
 \* 42012 42029: gap of unknown length  
 \* 42030 50909: contig of 8861 bp in length  
 \* 50911 50981: gap of unknown length  
 \* 50982 61359: contig of 10451 bp in length  
 \* 61360 61377: gap of unknown length  
 \* 61378 69256: contig of 7879 bp in length  
 \* 69257 69274: gap of unknown length  
 \* 69275 82266: contig of 12992 bp in length  
 \* 82267 82284: gap of unknown length  
 \* 82285 96734: contig of 14450 bp in length  
 \* 96735 96752: gap of unknown length  
 \* 96753 121042: contig of 24290 bp in length  
 \* 121043 121060: gap of unknown length  
 \* 121061 142804: contig of 21744 bp in length  
 \* 142805 142822: gap of unknown length  
 \* 142823 165414: contig of 22592 bp in length.

FEATURES  
 source  
 1. .165414  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NH0509E16"

BASE COUNT 41084 a 41694 c 40305 g 42043 t 288 others  
 ORIGIN

Query Match 1.7%; Score 46; DB 42; Length 165414;  
 Best Local Similarity 100.0%; Pred. NO. 2.5e-14;  
 Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689  
 |||||  
 Db 68785 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 68740

RESULT 28  
 AC009969/c  
 LOCUS  
 DEFINITION Homo sapiens chromosome unknown clone NH0418G17, WORKING DRAFT  
 SEQUENCE, in unordered pieces.

ACCESSION AC009969  
 VERSION AC009969.3 GI:6139226  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT  
 SOURCE human  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 TITLE 1 (bases 1 to 165445)  
 JOURNAL Sulston, J.E. and Waterston, R.  
 MEDLINE Toward a complete human genome sequence  
 REFERENCE 2 (bases 1 to 165445)  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 AUTHORS Waterston, R.  
 TITLE The sequence of Homo sapiens unknown clone NH0418G17  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 165445)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT  
 On Oct 29, 1999 this sequence version replaced gi:5923976.  
 SUBMITTED BY: WUGSC  
 TITLE Genome Sequencing Center  
 DEPARTMENT Department of Genetics  
 LOCATION Washington University  
 ST. LOUIS St. Louis MO 63108, USA  
 MAILTO:mailto:sapiens@watson.wustl.edu

NOTICE: This 'working draft' quality sequence may consist of several contigs from automated sequence assembly concatenated together. No attempt has been made to order or orient the contigs relative to one another correctly before concatenating. At each location in the sequence where contigs have been joined, several consecutive Ns may have been inserted.

The attached annotation was produced using a purely automated procedure.

The location of this clone is unknown.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

FEATURES  
 source  
 1. 165445  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="unknown"  
 /clone="NH0418G17"  
 /clone\_lib="unknown"  
 1. .972  
 /rpt\_family="Retroviral"  
 994. 1423  
 /rpt\_family="MER4-group"  
 1438. .1797  
 /rpt\_family="MaLR"  
 2998. .3497  
 /rpt\_family="MER4-group"  
 3861. 4238  
 /rpt\_family="MER21-group"  
 4891. 4938  
 /rpt\_family="(GGAA)n"  
 5536. .5885  
 /rpt\_family="MaLR"  
 5911. .5962  
 /rpt\_family="L1"  
 5963. 6266  
 /rpt\_family="Alu"  
 6267. 6690  
 /rpt\_family="L1"



```
repeat_region 6746..6796 /rpt_family="MER1_type"
repeat_region 7119..7216 /rpt_family="MaLR"
repeat_region 7373..7440 /rpt_family="MER1_type"
repeat_region 8177..8306 /rpt_family="MER1_type"
repeat_region 8919..9756 /rpt_family="L2"
repeat_region 9799..11055 /rpt_family="L1"
repeat_region 11062..11096 /rpt_family="L1"
repeat_region 11097..11374 /rpt_family="(GTTG)n"
repeat_region 11375..12677 /rpt_family="Alu"
repeat_region 13272..13558 /rpt_family="L1"
repeat_region 13954..14036 /rpt_family="Alu"
repeat_region 14037..14069 /rpt_family="L1"
repeat_region 14699..14821 /rpt_family="AT-rich"
repeat_region 15009..15137 /rpt_family="(TA)n"
repeat_region 15110..15459 /rpt_family="(TA)n"
repeat_region 15463..15565 /rpt_family="(TATATG)n"
repeat_region 15568..15741 /rpt_family="(TA)n"
repeat_region 15755..15915 /rpt_family="(TATAA)n"
repeat_region 16692..17574 /rpt_family="(TA)n"
repeat_region 17575..17827 /rpt_family="L1"
repeat_region 17828..18595 /rpt_family="Alu"
repeat_region 18596..19002 /rpt_family="L1"
repeat_region 19003..20453 /rpt_family="MaLR"
repeat_region 20484..20801 /rpt_family="L1"
repeat_region 20802..20829 /rpt_family="L1"
repeat_region 20830..21249 /rpt_family="(TAAAA)n"
repeat_region 21250..21545 /rpt_family="L1"
repeat_region 21546..22591 /rpt_family="Alu"
repeat_region 22601..23607 /rpt_family="L1"
repeat_region 23608..23696 /rpt_family="L1"
repeat_region 23698..24789 /rpt_family="Alu"
repeat_region 24790..24824 /rpt_family="L1"
repeat_region 24825..26046 /rpt_family="(CA)n"
repeat_region 26174..26540 /rpt_family="L1"
repeat_region 26601..28221 /rpt_family="MaLR"
repeat_region 28222..28520 /rpt_family="MaLR"
repeat_region 29411..29442 /rpt_family="MaLR"
```

```
repeat_region /rpt_family="(TCCC)n"
misc_feature 29438..29479 /rpt_family="CT-rich"
misc_feature 30423..30503 /note="pseudogene similar to PID:g3970874 (AB015344)
HRHFB2157 [Homo sapiens]"
misc_feature 30506..31558 /note="pseudogene similar to PID:g3970874 (AB015344)
HRHFB2157 [Homo sapiens]"
exon 31819..32103
repeat_region 32277..33564 /rpt_family="L2"
repeat_region 33838..34006 /rpt_family="MIR"
repeat_region 34246..34583 /rpt_family="MaLR"
repeat_region 34596..34784 /rpt_family="MER1_type"
repeat_region 35006..35126 /rpt_family="MER1_type"
repeat_region 35127..35247 /rpt_family="Alu"
repeat_region 35248..35315 /rpt_family="MER1_type"
repeat_region 35519..35668 /rpt_family="MIR"
repeat_region 35794..36077 /rpt_family="Alu"
repeat_region 36286..36497 /rpt_family="L2"
repeat_region 37034..37263 /rpt_family="Alu"
repeat_region 37655..37837 /rpt_family="Retroviral"
repeat_region 39386..39557 /rpt_family="MIR"
repeat_region 39639..39760 /rpt_family="L2"
repeat_region 39794..39864 /rpt_family="L2"
repeat_region 40988..41012 /rpt_family="(T)n"
repeat_region 41091..41348 /rpt_family="MIR"
repeat_region 41721..42106 /rpt_family="Retroviral"
repeat_region 43882..43999 /rpt_family="L2"
repeat_region 44623..44669 /rpt_family="(TG)n"
repeat_region 44756..44953
```

Query Match 1.7%; Score 46; DB 42; Length 165445;  
Best Local Similarity 100.0%; Pred. No. 2.5e-14;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2644 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689

Db 161469 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 161424

RESULT 29

AC012114/c  
LOCUS AC012114 168363 bp DNA HTG 20-OCT-1999  
DEFINITION Homo sapiens clone L\_H\_12, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 25  
unordered pieces.  
ACCESSION AC012114  
VERSION AC012114.1 GI:6088012  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

[illegible]

```
Center project name: L3777
Center clone name: 363.G.3
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142869 bases at least Q40
Consensus quality: 157748 bases at least Q30
Consensus quality: 164454 bases at least Q20
Insert size: 169168; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 675: contig of 675 bp in length
*      gap of unknown length
* 676 2714: contig of 2039 bp in length
*      gap of unknown length
* 2715 5561: contig of 2847 bp in length
*      gap of unknown length
* 5562 7526: contig of 1965 bp in length
*      gap of unknown length
* 7527 9378: contig of 1852 bp in length
*      gap of unknown length
* 9379 19384: contig of 10006 bp in length
*      gap of unknown length
* 19385 26444: contig of 7060 bp in length
*      gap of unknown length
* 26445 35340: contig of 8896 bp in length
*      gap of unknown length
* 35341 44797: contig of 9457 bp in length
*      gap of unknown length
* 44798 54146: contig of 9349 bp in length
*      gap of unknown length
* 54147 64634: contig of 10488 bp in length
*      gap of unknown length
* 64635 76415: contig of 11781 bp in length
*      gap of unknown length
* 76416 97235: contig of 20820 bp in length
*      gap of unknown length
* 97236 129656: contig of 32421 bp in length
*      gap of unknown length
* 129657 169168: contig of 39512 bp in length.
*      Location/Qualifiers
*          1..169168
*          /organism="Homo sapiens"
*          /db_xref="taxon:9606"
*          /chromosome="2"
*          /map="2"
*          /clone="RP11-363G3"
*          /clone_lib="RPC1-11 Human Male BAC"
*
* BASE COUNT 54711 a 28671 c 28866 g 56914 t 6 others
* ORIGIN
*
* Query Match 1..7%; Score 46; DB 44; Length 169168;
* Best Local Similarity 100.0%; Pred. No. 2.5e-14;
* Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
*      |||||||.....
* Db 19636 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 19681
*
* FEATURES
* Source
*
* BASE COUNT 342 bp mRNA PLN
* ORIGIN
*
* Query Match 1..7%; Score 46; DB 44; Length 169168;
* Best Local Similarity 100.0%; Pred. No. 2.5e-14;
* Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
*      |||||||.....
* Db 19636 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 19681
*
* RESULT 31
* SSV10781
* LOCUS 54711 a 342 bp mRNA PLN 14-OCT-1998
```

```
DEFINITION S.stapfianus PSD.8b mRNA.
ACCESSION Y10781
VERSION Y10781.1 GI:1808681
KEYWORDS desiccation tolerance.
SOURCE Sporobolus stapfianus.
ORGANISM Sporobolus stapfianus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Sporobolus.
REFERENCE 1 (bases 1 to 342)
AUTHORS Blomstedt,C.K., Gianello,R.D., Neale,A.D., Hamill,J.D. and
Gaff,D.F.
TITLE Isolation and characterization of cDNAs associated with the onset
of desiccation tolerance in the resurrection grass Sporobolus
stapfianus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 342)
AUTHORS Blomstedt,C.K.
DIRECT SUBMISSION
TITLE Submitted (20-JAN-1997) C.K. Blomstedt, Monash University, Genetics
And Developmental Biology, Wellington Road, Clayton, Melbourne,
Victoria 3168 AUSTRALIA
JOURNAL
FEATURES
source
1..342
/organism="Sporobolus stapfianus"
/variety="Gandoger"
/db_xref="taxon:56623"
/tissue_type="leaf"
/clone="psd.8b"
<1..60
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA71753.1"
/db_xref="GI:1808682"
/db_xref="SPTREMBL:O04817"
/translation="QKLYDDLTSGERNLEDEAR"
BASE COUNT 127 a 59 c 67 g 89 t
ORIGIN
*
* Query Match 1..7%; Score 45; DB 7; Length 342;
* Best Local Similarity 100.0%; Pred. No. 9.3e-14;
* Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* QY 2644 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2688
*      |||||||.....
* Db 298 CCCCCAAAAA.....AAAAAAAAAAAAAAAAAAAAAAAAAAAA 342
*
* RESULT 32
* AF005158
* LOCUS AF005158 1016 bp mRNA PLN 07-OCT-1998
* DEFINITION Arabidopsis thaliana MADS-box Protein (AGL24) mRNA, complete cds.
* ACCESSION AF005158
* VERSION AF005158.1 GI:3719214
* KEYWORDS thale cress.
* SOURCE Arabidopsis thaliana
* ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 1016)
AUTHORS Kohalmi,S.E., Schorr,P., Nowak,J., Reader,L.J. and Crosby,W.L.
TITLE The product of AGAMOUS interacts with select members of the
MADS-box family of proteins in Arabidopsis thaliana
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1016)
AUTHORS Kohalmi,S.E., Schorr,P., Nowak,J., Reader,L.J. and Crosby,W.L.
DIRECT SUBMISSION
TITLE Submitted (22-MAY-1997) Gene Expression Group, NRC Plant
Biotechnology Institute, 110 Gymnasium Place, Saskatoon, SK S7N
0W9, Canada
```

```

FEATURES
  source
    Location/Qualifiers
      1..1016
        /organism="Arabidopsis thaliana"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /map="Linked to m600 and PG11"
        /chromosome="4"
      1..1016
        /gene="AGL24"
      96..758
        /gene="AGL24"
        /codon_start=1
        /product="MADS-box Protein"
        /protein_id="AAC63139.1"
        /db_xref="GI:3719215"
        /translation="MAREKIRIKKIDNTAROVTFESKRRGIFKKADELSVLCADVA
        LIISATOKLEFESSSRNDILGRYSLHASNINKLMODPPSHLENCNLSRLSKEVE
        DTKOLKRGEDDGLMLELORLEKLEGLSRVSEKRGECVMSOIFSLKRGSEL
        VDENKRLRKLETLERAKLTTLKEALETESVTINVSYSYDGTPLDDSDTSLSKLGIPS
        WE"
BASE COUNT      352 a 189 c 244 g 231 t
ORIGIN
Query Match      1.7%; Score 45; DB 8; Length 1016;
Best Local Similarity 100.0%; Pred. No. 9.3e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||
Db 953 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 997

RESULT 33
LOCUS      AF012536      1180 bp      mRNA      PRI      21-AUG-1997
DEFINITION Homo sapiens decoy receptor 1 (Dcrl) mRNA, complete cds.
ACCESSION      AF012536
VERSION      AF012536.1 GI:2338421
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  AUTHORS      Sheridan,J.P., Marsters,S.A., Pitti,R.M., Gurney,A., Skubatch,M.,
      Baldwin,D., Ramakrishnan,L., Gray,C.L., Baker,K., Wood,W.I.,
      Goddard,A.D., Godowski,P. and Ashkenazi,A.
  TITLE      Control of TRAIL-induced apoptosis by a family of signaling and
      decoy receptors
  JOURNAL      Science 277 (5327), 818-821 (1997)
MEDLINE      97390509
REFERENCE
  AUTHORS      Sheridan,J.P., Marsters,S.A., Pitti,R.M., Gurney,A., Baldwin,D.,
      Ramakrishnan,L., Gray,C.L., Baker,K., Wood,W.I., Goddard,A.D.,
      Godowski,P. and Ashkenazi,A.
  TITLE      Direct Submission
  JOURNAL      Submitted (06-JUL-1997) Molecular Oncology, Genentech, 1 DNA Way,
      South San Francisco, CA 94080, USA
FEATURES
  source
    Location/Qualifiers
      1..1180
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
      1..1180
        /gene="Dcrl"
      193..972
        /gene="Dcrl"
        /note="tumor necrosis factor receptor family member;
        inhibits apoptosis induction by TRAIL/Apo2L"
        /codon_start=1
        /product="decoy receptor 1"
        /protein_id="AAB67104.1"
        /db_xref="GI:2338422"

gene
  polyA_site 954 a 780 c 702 g 669 t
CDS
  ORIGIN
Query Match      1.7%; Score 45; DB 10; Length 3105;
Best Local Similarity 100.0%; Pred. No. 9.2e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT      338 a 326 c 298 g 218 t
ORIGIN
/translation="MARIPKTLKVVVAVLLPVLAYSATTARQEEVPOQTVAPOQO
RHSEKGECPAGSHRSHTGACNPCTEGVDYTNASNNPSCFCPTCKCKDOKHKSCT
MTRDTVCQCKEGTFERNENSPCMGRKSCPCSGEVOVSNCTSMDDIQCVEEGANATVE
TPAAEETMTSPGTPAPAAEETMTSPGTPAPAAEETMTSPGTPAPAAEETMTSPG
TPAAEETMTSPGTPASSHYLSCTIIGIIVLILVIFV"
Query Match      1.7%; Score 45; DB 10; Length 1180;
Best Local Similarity 100.0%; Pred. No. 9.3e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||
Db 1108 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1152

RESULT 34
LOCUS      HSM800649      3105 bp      mRNA      PRI      23-JUN-1999
DEFINITION Homo sapiens mRNA; CDNA DKFZp434J193 (from clone DKFZp434J193).
ACCESSION      AL080137
VERSION      AL080137.1 GI:5262579
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 3105)
  AUTHORS      Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE      Direct Submission
  JOURNAL      Submitted (22-JUN-1999) MIPS, Am Klopferspitz 18a D-82152
      Martinsried, GERMANY
  COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
      Research Center (DKFZ); sequenced by AGOWA within the CDNA
      sequencing consortium of the German Genome Project. This clone is
      available at the RZPD in Berlin.
      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
  source
    Location/Qualifiers
      1..3105
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="DKFZp434J193"
        /clone_lib="434 (synonym: htes3). Vector pSport1; host
        DH10B; sites NotI + SalI"
        /dev_stage="adult"
        /tissue_type="testis"
      1..862
        /gene="DKFZp434J193"
        /note="weak similarity to neurofilament subunit"
        /codon_start=2
        /product="hypothetical protein"
        /protein_id="CAB45731.1"
        /db_xref="GI:5262580"
        /translation="HILTKLNFHLKVKLEIOWGIPIRAKRSQETAAPENISTOKS
        LESLNHOGETLLLOELPTPDTPAPNPGVHLKEOLANDLKAVQOKOSKAVPOG
        SAHSVSKISQSGDMTEAHNPCVQVEASVNNPSLEECGPEQSPSKSDPAHPMLA
        SHAHDEETRAADYREGDAGFGSRGERRPAEDQRPAGMLPNTPTPRGWSRSF
        HLADPCQHSQPHHPQFKLPQLPFRVPFGESEKDLQDSQTKLTVILLEPATIPENTQTV
        VPQLHRVLSL"
      2992
      polyA_site
BASE COUNT      954 a 780 c 702 g 669 t
ORIGIN
Query Match      1.7%; Score 45; DB 10; Length 3105;
Best Local Similarity 100.0%; Pred. No. 9.2e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Qy 2645 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||
Db 2989 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3033

RESULT 35
AF106934 3220 bp mRNA PRI 19-MAY-1999
LOCUS Homo sapiens vitamin D receptor-interacting protein (DRIP92) mRNA,
DEFINITION complete cds.
ACCESSION AF106934
VERSION AF106934.1 GI:4868007
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Rachez,C., Lemon,B.D., Suldan,Z., Bromleigh,V., Gamble,M.,
Naar,A.M., Erdjument-Bromage,H., Tempst,P. and Freedman,L.P.
TITLE Ligand-dependent transcription activation by nuclear receptors
requires the DRIP complex
JOURNAL Nature 398 (6730), 824-828 (1999)
MEDLINE 99249345
AUTHORS Gamble,M., Rachez,C. and Freedman,L.P.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1998) Cell Biology, MSKCC, 1275 York Avenue, New
York, NY 10021, USA
FEATURES
source Location/Qualifiers
1..3220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="B-cell"
1..3220
/gene="DRIP92"
/gene="DRIP92"
/Note="92 kDa component of the DRIP coactivator complex
for the vitamin D3 receptor."
/codon_start=1
/product="vitamin D receptor-interacting protein"
/protein_id="AAD31087.1"
/db_xref="GI:4868008"
/translation="MCDLRRAAGMDLAYVCEWEKWSKSTHCPSVPLACAWSCRL
IAFTDLRSDDODLTRMTHILDTEHPDLHSIPSEHHEAITCLEWDQSGSRLLSADAD
GOIKCWSADHLANSWESSVGLVGDPIVALSHLNGVKIALHVERSGASSFGEKES
RYKFSPLTFGKPMEGNIATVSLVTSLLKPSGQVLSTESLCRLRARVALADI
AFTGGNIIVATADGSSASPVQFYKVCVSVSEKCRIDTEILPSLFMRCTTDLNRKDK
FPATITHLFLARDMSEQVILLCASSQTSIVCEWSLRKEGLPVNNIFQOISFVWDGSP
TILKWRILSATNDLRSVALPKLPISLTNTDLKVASDQTFYPLGLALAFHDGSYH
IVHRLSLOTMAVYSSAAPRPVDEPAMKRPRTAGFAVHLKAMQLSWTSLAVGIDSHG
KLVLRLSPSGMHPLEGLARHLHLLFLEYCMVTGYDMDWDLHVPQSMVQSLVEKLH
EYRTQALQGVLTSTRILAMKASLCKLSPCTVTRVCDYHFKLFLLAISLTKSLLRP
HEFNTPKDPSGRLLEICIKITDVIDKVMNLKTEEFVDMTHCRKSSCSGWAITS
CTCWPAPYTPAPRSPAPRSPPPPTATSDQSMSLFLRLKLMCCRCDEGASE
SLGMLRELVVIRWGLLPSGLPYTATSDQSMSLFLRLKLMCCRCDEGASE
PDEALVDECCLLPSOLLTSLDMLPASGLVSRLOPKPLQFGRAPTLPQSAATIQ
LDGLARAPGQPKIDHLRLHLGLGACPTCECKACGCTGCTMLKSPNRTAVKQWEQWFI
KNLCGLGLWVRVPLSPY"
BASE COUNT 596 a 1125 c 955 g 544 t
ORIGIN

Query Match 1.7%; Score 45; DB 40; Length 3220;
Best Local Similarity 100.0%; Pred. No. 9.2e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 45; Conservative 0;

Qy 2645 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||
Db 3166 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3210

RESULT 36
AR068182 3350 bp DNA PAT 29-SEP-1999
LOCUS Sequence 1 from patent US 5853733.
DEFINITION AR068182
ACCESSION AR068182
VERSION AR068182.1 GI:6000389
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3350)
AUTHORS Cochran,M.D. and Macdonald,R.D.
TITLE Recombinant herpesvirus of turkeys and uses thereof
JOURNAL Patent: US 5853733-A 1 29-DEC-1998;
FEATURES
source Location/Qualifiers
1..3350
/organism="unknown"
BASE COUNT 941 a 912 c 859 g 638 t
ORIGIN

Query Match 1.7%; Score 45; DB 5; Length 3350;
Best Local Similarity 100.0%; Pred. No. 9.2e-14; Mismatches 0; Indels 0; Gaps 0;
Matches 45; Conservative 0;

Qy 2645 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
      |||
Db 3246 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3290

RESULT 37
S73375 3582 bp mRNA MAM 28-FEB-1995
LOCUS p25/p35=Cyclin-dependent kinase 5 activator [cattle, brain, mRNA,
DEFINITION 3582 nt].
ACCESSION S73375
VERSION S73375.1 GI:685111
KEYWORDS Bos taurus brain.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
AUTHORS Lew,J., Huang,Q.Q., Qi,Z., Winkfein,R.J., Aebersold,R., Hunt,T. and
Wang,J.H.
TITLE A brain-specific activator of cyclin-dependent kinase 5
JOURNAL Nature 371 (6496), 423-426 (1994)
MEDLINE 94376896
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsg 155522] from the original journal article.
This sequence comes from Fig. 1.
FEATURES
source Location/Qualifiers
1..3582
/organism="Bos taurus"
/db_xref="taxon:9913"
464..1387
/gene="p25/p35"
/Note="Cyclin-dependent kinase 5 activator, Cdk5
activator, p25/p35"
464..1387
/gene="p25/p35"
/Note="p35 subunit readily degraded to p25 subunit is also
active. Method: conceptual translation with partial
peptide sequencing. This sequence comes from Fig. 1; Cdk5
activator: p25/p35"
/codon_start=1
/product="Cyclin-dependent kinase 5 activator"
/protein_id="AAB31984.1"
/db_xref="GI:685112"
/translation="MGTVLISLSPSYRYKATLFEDGAATVGHYTAQVNSKNAKDKLKRH
SIISVLPMKRIYAVSAKKNSKVQPNSSYQNNITHLNENLKSLSCANLSTFAQPP
PAQPPAPPASQLSGSQTGYSSSVKKAPHPAVSSAGTPKRVITVQATSELLRCLGELC
```

RCYRLKHLSPDPVWLRLSVDRSLLOGMDOGFIPTANVVELYMLCRDVISSEVGS  
 DHELAVLLTCLYLSVYMGNEISVPLAPFLVESCKEAFWDRCLSVINLMSSKMLQIN  
 ADPHFTQVFSDLKNESQDEKRLLLGLDR  
 BASE COUNT 694 a 1113 c 1084 g 691 t  
 ORIGIN

Query Match 1.7%; Score 45; DB 3; Length 3582;

Best Local Similarity 100.0%; Pred. No. 9.2e-14; Mismatches 0; Indels 0; Gaps 0;

Qy 2645 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689  
 I|||||

Db 3538 CCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3582

## RESULT 38

AC008696/c

LOCUS

DEFINITION

Homo sapiens chromosome 5 clone CIT978SKB\_70D19, \*\*\* SEQUENCING IN

AC008696

VERSION

AC008696.1

KEYWORDS

HTG; HTGS\_PHASE1.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 41684)

DOE Joint Genome Institute.

TITLE

Sequencing of Human Chromosome 5

JOURNAL

Unpublished

2 (bases 1 to 41684)

DOE Joint Genome Institute.

Direct Submission

Submitted (03-AUG-1999)

Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

www.jgi.doe.gov.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 83 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1

\* 661: contig of 661 bp in length

\* gap of unknown length

\* 662 1321: contig of 660 bp in length

\* gap of unknown length

\* 1322 1605: contig of 284 bp in length

\* gap of unknown length

\* 1606 2158: contig of 553 bp in length

\* gap of unknown length

\* 2159 2814: contig of 656 bp in length

\* gap of unknown length

\* 2815 2913: contig of 99 bp in length

\* gap of unknown length

\* 2914 2976: contig of 63 bp in length

\* gap of unknown length

\* 2977 3046: contig of 70 bp in length

\* gap of unknown length

\* 3047 4099: contig of 1053 bp in length

\* gap of unknown length

\* 4100 5221: contig of 1122 bp in length

\* gap of unknown length

\* 5222 5877: contig of 656 bp in length

\* gap of unknown length

\* 5878 6061: contig of 184 bp in length

\* gap of unknown length

\* 6062 6178: contig of 117 bp in length

\* gap of unknown length

\* 6179 6636: contig of 458 bp in length

6637 6978: contig of 242 bp in length  
 gap of unknown length  
 6879 6940: contig of 62 bp in length  
 gap of unknown length  
 6941 7030: contig of 90 bp in length  
 gap of unknown length  
 7031 8041: contig of 1011 bp in length  
 gap of unknown length  
 8042 8098: contig of 57 bp in length  
 gap of unknown length  
 8099 8423: contig of 325 bp in length  
 gap of unknown length  
 8424 9390: contig of 967 bp in length  
 gap of unknown length  
 9391 9978: contig of 588 bp in length  
 gap of unknown length  
 9979 10631: contig of 653 bp in length  
 gap of unknown length  
 10632 11504: contig of 873 bp in length  
 gap of unknown length  
 11505 11550: contig of 46 bp in length  
 gap of unknown length  
 11551 11779: contig of 229 bp in length  
 gap of unknown length  
 11780 11870: contig of 91 bp in length  
 gap of unknown length  
 11871 11941: contig of 71 bp in length  
 gap of unknown length  
 11942 12029: contig of 88 bp in length  
 gap of unknown length  
 12030 12122: contig of 93 bp in length  
 gap of unknown length  
 12123 12224: contig of 102 bp in length  
 gap of unknown length  
 12225 13275: contig of 1051 bp in length  
 gap of unknown length  
 13276 13390: contig of 115 bp in length  
 gap of unknown length  
 13391 14045: contig of 655 bp in length  
 gap of unknown length  
 14046 14963: contig of 918 bp in length  
 gap of unknown length  
 14964 15144: contig of 181 bp in length  
 gap of unknown length  
 15145 15812: contig of 668 bp in length  
 gap of unknown length  
 15813 15889: contig of 77 bp in length  
 gap of unknown length  
 15890 16083: contig of 194 bp in length  
 gap of unknown length  
 16084 16175: contig of 92 bp in length  
 gap of unknown length  
 16176 16773: contig of 598 bp in length  
 gap of unknown length  
 16774 17972: contig of 1199 bp in length  
 gap of unknown length  
 17973 18078: contig of 106 bp in length  
 gap of unknown length  
 18079 19129: contig of 1051 bp in length  
 gap of unknown length  
 19130 19690: contig of 561 bp in length  
 gap of unknown length  
 19691 19772: contig of 82 bp in length  
 gap of unknown length  
 19773 20041: contig of 269 bp in length  
 gap of unknown length  
 20042 20398: contig of 357 bp in length  
 gap of unknown length  
 20399 21269: contig of 871 bp in length  
 gap of unknown length  
 21270 21399: contig of 130 bp in length  
 gap of unknown length







details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES  
source  
1..111079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q31"  
/clone\_lib="RPCI-3"  
/clone="RP3-419C19"  
BASE COUNT 35388 a 21441 c 21433 g 32817 t  
ORIGIN

Query Match 1.7%; Score 45; DB 11; Length 111079;  
Best Local Similarity 100.0%; Pred. No. 9e-14;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCAA 2689  
|||||  
Db 13113 CCCCAA 13069

RESULT 41  
AC011414/c  
LOCUS Homo sapiens chromosome 5 clone C1978SKB\_83D3, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 11 unordered pieces.  
ACCESSION AC011414  
VERSION AC011414.1 GI:6013528  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 125260)  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
AUTHORS DOE Joint Genome Institute.  
REFERENCE 2 (bases 1 to 125260)  
DOE Joint Genome Institute.  
DIRECT SUBMISSION  
SUBMITTED (06-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
www.jgi.doe.gov.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 660: contig of 660 bp in length  
gap of unknown length  
\* 661 1339: contig of 679 bp in length  
gap of unknown length  
\* 1340 2021: contig of 682 bp in length  
gap of unknown length  
\* 2022 5034: contig of 3013 bp in length  
gap of unknown length  
\* 5035 10078: contig of 5044 bp in length  
gap of unknown length  
\* 10079 21131: contig of 11053 bp in length  
gap of unknown length  
\* 21132 33664: contig of 12533 bp in length  
gap of unknown length  
\* 33665 46169: contig of 12505 bp in length  
gap of unknown length  
\* 46170 63481: contig of 17312 bp in length  
gap of unknown length  
\* 63482 92267: contig of 28786 bp in length  
gap of unknown length  
\* 92268 125260: contig of 32993 bp in length.

FEATURES  
source  
1..125260  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="C1978SKB\_83D3"  
BASE COUNT 40541 a 24231 c 24261 g 36163 t 64 others  
ORIGIN

Query Match 1.7%; Score 45; DB 41; Length 125260;  
Best Local Similarity 100.0%; Pred. No. 9e-14;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCAA 2689  
|||||  
Db 11448 CCCCAA 11404

RESULT 42  
HSDJ319M7/c  
LOCUS Homo sapiens chromosome 6 clone RP1-319M7 map p21.1-21.3, \*\*\*  
DEFINITION SEQUENCING IN PROGRESS \*\*\*, in unordered pieces.  
ACCESSION AL079341  
VERSION AL079341.14 GI:6434655  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 128208)  
AUTHORS Corby, N.  
TITLE Direct Submission  
JOURNAL Submitted (30-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Nov 15, 1999 this sequence version replaced gi:6136977.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in  
progress and the release of this data is based on the understanding  
that the sequence may change as work continues. The sequence may  
be contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments. Contig\_ID: 01957 Length: 128208bp.  
\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1..128208  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP1-319M7"  
/clone\_lib="RPCI-1"  
/map="p21.1-21.3"  
BASE COUNT 36143 a 23938 c 26492 g 41630 t 5 others  
ORIGIN

Query Match 1.7%; Score 45; DB 33; Length 128208;  
Best Local Similarity 100.0%; Pred. No. 9e-14;  
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2645 CCCCAA 2689  
|||||  
Db 101579 CCCCAA 101535

RESULT 43  
AC011262  
LOCUS AC011262 145346 bp DNA HTG 19-NOV-1999

```

DEFINITION Homo sapiens clone RP11-2P11, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011262
VERSION AC011262.2 GI:6454026
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145346)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-2P11
Unpublished
2 (bases 1 to 145346)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehocky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severi,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 19, 1999 this sequence version replaced gi:6012128.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L2689
Center clone name: 2_P11
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 97751 bases at least Q40
Consensus quality: 121236 bases at least Q30
Consensus quality: 136150 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 145346; sum-of-ctngs
Quality coverage: 3.1 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1091: contig of 1091 bp in length
gap of unknown length
* 1092 2456: contig of 1365 bp in length
gap of unknown length
* 2457 3980: contig of 1524 bp in length
gap of unknown length
* 3981 6275: contig of 2295 bp in length
gap of unknown length
* 6276 8952: contig of 2677 bp in length
gap of unknown length

```

```

* 8953 11694: contig of 2742 bp in length
gap of unknown length
* 11695 12820: contig of 1126 bp in length
gap of unknown length
* 12821 15352: contig of 2532 bp in length
gap of unknown length
* 15353 19536: contig of 4184 bp in length
gap of unknown length
* 19537 23248: contig of 3712 bp in length
gap of unknown length
* 23249 28780: contig of 5532 bp in length
gap of unknown length
* 28781 32808: contig of 4028 bp in length
gap of unknown length
* 32809 37816: contig of 5008 bp in length
gap of unknown length
* 37817 43068: contig of 5252 bp in length
gap of unknown length
* 43069 48045: contig of 4977 bp in length
gap of unknown length
* 48046 56867: contig of 8822 bp in length
gap of unknown length
* 56868 67910: contig of 11043 bp in length
gap of unknown length
* 67911 81484: contig of 13574 bp in length
gap of unknown length
* 81485 99620: contig of 18136 bp in length
gap of unknown length
* 99621 117196: contig of 17576 bp in length
gap of unknown length
* 117197 145346: contig of 28150 bp in length.
Location/Qualifiers
1. 145346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-2P11"
/clone.lib="RPC1-11 Human Male BAC"
BASE COUNT 47197 a 25901 c 27079 g 45152 t 17 others
ORIGIN

Query Match 1-7% Score 45; DB 43; Length 145346;
Best Local Similarity 100.0%; Pred. No. 9e-14; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 0;

Qy 2645 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689
|||||
Db 99962 CCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 100006

RESULT 44
AC015500/c
LOCUS Homo sapiens clone RP11-21G19, WORKING DRAFT SEQUENCE, 10 unordered
DEFINITION AC015500 156730 bp DNA HTG 09-DEC-1999
pieces.
ACCESSION AC015500
VERSION AC015500.2 GI:6553985
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156730)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-21G19
Unpublished
2 (bases 1 to 156730)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

```



```
* 28595 31448: contig of 2854 bp in length
*      gap of unknown length
* 31449 35732: contig of 4284 bp in length
*      gap of unknown length
* 35733 40093: contig of 4361 bp in length
*      gap of unknown length
* 40094 44260: contig of 4167 bp in length
*      gap of unknown length
* 44261 49519: contig of 5259 bp in length
*      gap of unknown length
* 49520 55034: contig of 5515 bp in length
*      gap of unknown length
* 55035 63390: contig of 8356 bp in length
*      gap of unknown length
* 63391 70014: contig of 6624 bp in length
*      gap of unknown length
* 70015 75415: contig of 5401 bp in length
*      gap of unknown length
* 75416 82425: contig of 7010 bp in length
*      gap of unknown length
* 82426 90586: contig of 8161 bp in length
*      gap of unknown length
* 90587 97723: contig of 7137 bp in length
*      gap of unknown length
* 97724 106002: contig of 8279 bp in length
*      gap of unknown length
* 106003 115333: contig of 9331 bp in length
*      gap of unknown length
* 115334 125768: contig of 10435 bp in length
*      gap of unknown length
* 125769 140335: contig of 14567 bp in length
*      gap of unknown length
* 140336 155326: contig of 14991 bp in length
*      gap of unknown length
* 155327 162504: contig of 7178 bp in length.
```

FEATURES

```
source
1..162504
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="397_P_10"
/clone_lib="RFCl-11 Human Male BAC"
BASE COUNT 51669 a 29955 c 29241 g 50694 t 945 others
ORIGIN
```

```
Query Match 1.7% Score 45; DB 42; Length 162504;
Best Local Similarity 100.0%; Pred. No. 9e-14;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2644 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2688
      |
Db 39562 CCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 39518
```

Search completed: April 6, 2000, 14:55:06  
Job time: 27491 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:17:39 ; Search time 6421.81 Seconds  
(without alignments)  
1338.160 Million cell updates/sec

Title: US-09-090-672B-6  
Perfect score: 2276  
Sequence: 1 CTGAACTGGGAGTCAGGTGG.....AAAAAAAAAAAAAAAAAAAA 2276

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

EST.\*  
1: em\_est1.\*  
2: em\_est2.\*  
3: em\_est3.\*  
4: em\_est4.\*  
5: em\_est5.\*  
6: em\_est6.\*  
7: em\_est7.\*  
8: em\_est8.\*  
9: em\_est9.\*  
10: em\_est10.\*  
11: em\_est11.\*  
12: em\_est12.\*  
13: em\_est13.\*  
14: em\_est14.\*  
15: em\_est15.\*  
16: em\_est16.\*  
17: em\_est17.\*  
18: em\_est18.\*  
19: em\_est19.\*  
20: gb\_est1.\*  
21: gb\_est2.\*  
22: gb\_est3.\*  
23: gb\_est4.\*  
24: gb\_est5.\*  
25: gb\_est6.\*  
26: gb\_est7.\*  
27: gb\_est8.\*  
28: gb\_est9.\*  
29: gb\_est10.\*  
30: gb\_est11.\*  
31: gb\_est12.\*  
32: gb\_est13.\*  
33: gb\_est14.\*  
34: gb\_est15.\*  
35: gb\_est16.\*  
36: gb\_est17.\*  
37: gb\_est18.\*  
38: gb\_est19.\*  
39: gb\_est20.\*  
40: gb\_est21.\*  
41: gb\_est22.\*  
42: gb\_est23.\*  
43: gb\_est24.\*

44: gb\_est25.\*  
45: gb\_est26.\*  
46: gb\_est27.\*  
47: gb\_est28.\*  
48: gb\_est29.\*  
49: gb\_est30.\*  
50: gb\_est31.\*  
51: gb\_est32.\*  
52: em\_est20.\*  
53: em\_est21.\*  
54: em\_est22.\*  
55: em\_est23.\*  
56: em\_est24.\*  
57: em\_est25.\*  
58: em\_est26.\*  
59: gb\_est33.\*  
60: gb\_est34.\*  
61: gb\_est35.\*  
62: gb\_est36.\*  
63: gb\_est37.\*  
64: gb\_est38.\*  
65: em\_est27.\*  
66: em\_est28.\*  
67: em\_est29.\*  
68: em\_est30.\*  
69: gb\_est39.\*  
70: gb\_est40.\*  
71: gb\_est41.\*  
72: gb\_est42.\*  
73: gb\_est43.\*  
74: gb\_est44.\*  
75: em\_est31.\*  
76: em\_est32.\*  
77: em\_est33.\*  
78: em\_est34.\*  
79: gb\_gss1.\*  
80: gb\_gss2.\*  
81: gb\_gss3.\*  
82: gb\_gss4.\*  
83: em\_gss1.\*  
84: em\_gss2.\*  
85: em\_gss3.\*  
86: em\_gss4.\*  
87: gb\_gss5.\*  
88: gb\_gss6.\*  
89: gb\_gss7.\*  
90: gb\_gss8.\*  
91: gb\_gss9.\*  
92: em\_gss5.\*  
93: em\_gss6.\*  
94: em\_gss7.\*  
95: em\_gss8.\*  
96: em\_gss9.\*  
97: em\_gss10.\*  
98: em\_gss11.\*  
99: gb\_gss10.\*  
100: gb\_gss11.\*  
101: em\_gss12.\*  
102: gb\_gss12.\*  
103: gb\_gss13.\*  
104: gb\_gss14.\*  
105: gb\_gss15.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Query	No.	Score	Match	Length	DB	ID	Description
c	1	453	19.9	464	49	Ar660113	Ar660113 ws6902.x	



AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 18, 1998 this sequence version replaced gi:3137760.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 463.

## FEATURES

source

Location/Qualifiers  
1..478  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2345468"  
/clone\_lib="Soares\_NHLTh"  
/dev\_stage="fetal"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: thymus, pooled; Vector: pT73D-Pac  
(Pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
TGTTACCAATCGAAGTCGAGCGCGCAAGCTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 173 a 93 c 59 g 153 t

ORIGIN

Query Match 19.98; Score 450; DB 50; Length 478;  
Best Local Similarity 100.0%; Pred. No. 6.4e-185;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1810 TTGTGTATGGTGTGTGTAACCTGTTTAAATTAAGTGAACCTGCTGAGACAGAGCTAT 1869

Db 450 TTGTGTATGGTGTGTGTAACCTGTTTAAATTAAGTGAACCTGCTGAGACAGAGCTAT 391

Qy 1870 TCTCCATGTAAGCAAGCTGATTTCTGAGCAATTTAATGATGCCGCTGGAGTACA 1929

Db 390 TCTCCATGTAAGCAAGCTGATTTCTGAGCAATTTAATGATGCCGCTGGAGTACA 331

Qy 1930 AAAGTGGAGTGGCTGAGTAATGATGATGGTGTGTTTACCATTCTTTGAGGTAAAG 1989

Db 330 AAAGTGGAGTGGCTGAGTAATGATGATGGTGTGTTTACCATTCTTTGAGGTAAAG 271

Qy 1990 CATCATGTAACCTGTAAGGAATTTAAATCCTACTTTTCATAATAAGTTGCATAGTT 2049

Db 270 CATCATGTAACCTGTAAGGAATTTAAATCCTACTTTTCATAATAAGTTGCATAGTT 211

Qy 2050 TAATAATTTTAAATATATGCTTGAGTTTAAATTTAAATGTAAGCGTAACCTTTAACT 2109

Db 210 TAATAATTTTAAATATATGCTTGAGTTTAAATTTAAATGTAAGCGTAACCTTTAACT 151

Qy 2110 CTATAATGTTTCTTCTGGAATTAATCCTCAACATATGAATTAATGTTGCTGCTT 2169

Db 150 CTATAATGTTTCTTCTGGAATTAATCCTCAACATATGAATTAATGTTGCTGCTT 91

Qy 2170 CCAAGAGCCTTTTGTGAAAAAGCTTTTGTGAATCATCAAGCTTTTCACTTTAAAT 2229

Db 90 CCAAGAGCCTTTTGTGAAAAAGCTTTTGTGAATCATCAAGCTTTTCACTTTAAAT 31

Qy 2230 AAAGTGTGTAAGCTTTTATTAATAAAA 2259

Db 30 AAAGTGTGTAAGCTTTTATTAATAAAA 1

RESULT 3

AW183584/c

LOCUS AW183584 494 bp mRNA EST 18-NOV-1999  
DEFINITION xj77g10.x1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
IMAGE:2663298 3', mRNA sequence.  
ACCESSION AW183584  
VERSION AW183584.1 GI:6452098  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 494)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Dec 20, 1995 this sequence version replaced gi:1135156.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 459.

## FEATURES

source

Location/Qualifiers

1..494

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2663298"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NbHL19W, testis NHT, and B-cell  
NCI-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 176 a 100 c 64 g 154 t

ORIGIN

Query Match 19.5%; Score 443; DB 74; Length 494;  
Best Local Similarity 99.8%; Pred. No. 6.8e-182;  
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1755 TGGGCCATATGAAAGGCTAAGCTTCACTGTAAATAATAACTGGGAATTCCTGGTTGTG 1814

Db 494 TGGGCCATATGAAAGGCTAAGCTTCACTGTAAATAATAACTGGGAATTCCTGGATTGTG 435

Qy 1815 TATGGGTCTTGGTGAACCTGTTTAAATTAAGTGAACCTGCTGAGAGACAGAGCTATCTCC 1874

Db 434 TATGGGTCTTGGTGAACCTGTTTAAATTAAGTGAACCTGCTGAGAGACAGAGCTATCTCC 375

Qy 1875 ATGTAAGTGGCAAGACCTGATTCTGAGCAATTAATATGATGCCGCTGGAGTACAAAGT 1934

Db 374 ATGTAAGTGGCAAGACCTGATTCTGAGCAATTAATATGATGCCGCTGGAGTACAAAGT 315

Qy 1935 GGAGTGTGGCCTGAGTAATGCAATTAATGGTGGTTTACCATTCTTGAGGTAAAGCATCA 1994

Db 314 GGAGTGTGGCCTGAGTAATGCAATTAATGGTGGTTTACCATTCTTGAGGTAAAGCATCA 255

Qy 1995 CATGAACCTGTAAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTCATAGGTTTAAATA 2054

Db 254 CATGAACCTGTAAAGGAATTTAAAAATCCTACTTTTCATAATAAGTTCATAGGTTTAAATA 195

Qy 2055 ATTTTAAATATATGGCTTGAGTTTAAATTTGTAATAGCGGTAACTATTTTAACTCTATA 2114

Db 194 ATTTTAAATATATGGCTTGAGTTTAAATTTGTAATAGCGGTAACTATTTTAACTCTATA 135

```

QY 2115 ATGTGTTTCATTCGGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTTCCAAG 2174
|||||
Db 134 ATGTGTTTCATTCGGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTTCCAAG 75
|||||
QY 2175 ACCCTTTTTCAGAAAAAGCTTTTTCGAATCATCAAGTCTTTTCACATTTAAATAAAGT 2234
|||||
Db 74 ACCCTTTTTCAGAAAAAGCTTTTTCGAATCATCAAGTCTTTTCACATTTAAATAAAGT 15
|||||
QY 2235 GTTTGAAAGCTTTA 2248
|||||
Db 14 GTTTGAAAGCTTTA 1
|||||

RESULT 4
AI338791/c 493 bp mRNA EST 13-FEB-1999
LOCUS qt53f04.x1 Soares_fetal_lung_NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:1951711 3', mRNA sequence.
ACCESSION AI338791
VERSION AI338791
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 493)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 244 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 469.
Location/Qualifiers
1. 493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE.1951711"
/clone_lib="Soares_fetal_lung_NbHL19W"
/dev_stage="19 weeks"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; lstr
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTCAAGTGGGAGCGCGCAATTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."
BASE COUNT 175 a 97 c 64 g 157 t
ORIGIN

Query Match 19.3%; Score 439; DB 45; Length 493;
Best Local Similarity 100.0%; Pred. No. 3.7e-180;
Matches 439; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1810 TTGTGTATGGGTGGTGAACCTGGTGTATAGTGAACCTGCTGAGAGACAGAGCTAT 1869
|||||
Db 442 TTGTGTATGGGTGGTGAACCTGGTGTATAGTGAACCTGCTGAGAGACAGAGCTAT 383
|||||
QY 1870 TCTCCATGTACTGGCAAGACCTGATTTCTGAGCATTTAATATGGATGCCGGGAGATACA 1929
|||||

```

```

Db 382 TCTCCATGTACTGGCAAGACCTGATTTCTGAGCATTTAATATGGATGCCGTGGGAGTACA 323
QY 1930 AAAGTGGAGTGGCCCTGAGTAATGATGATGATGGTGGTTTACCATTTCTTGGAGGTAAAG 1989
|||||
Db 322 AAAGTGGAGTGGCCCTGAGTAATGATGATGATGGTGGTTTACCATTTCTTGGAGGTAAAG 263
|||||
QY 1990 CATCATGAACTGTGAAAGAAATTTAAATCCTACTTCTCATATAAGTTCATAGGTT 2049
|||||
Db 262 CATCATGAACTGTGAAAGAAATTTAAATCCTACTTCTCATATAAGTTCATAGGTT 203
|||||
QY 2050 TAATAATTTTAAATATATGCTTGGCTTGAATTTAAATGTAATAGCGCTAACTAATTTAACT 2109
|||||
Db 202 TAATAATTTTAAATATATGCTTGGCTTGAATTTAAATGTAATAGCGCTAACTAATTTAACT 143
|||||
QY 2110 CTATAATGCTTCATCTCGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTT 2169
|||||
Db 142 CTATAATGCTTCATCTCGAATAATCCTAAACATATGAATTAATGTTTGCATGTTCACTT 83
|||||
QY 2170 CCAAGAGCCCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAAAT 2229
|||||
Db 82 CCAAGAGCCCTTTTGTGAAAAAAGCTTTTGTGAATCATCAAGTCTTTCACATTTAAAT 23
|||||
QY 2230 AAAGTCTTTGAAAGCTTTA 2248
|||||
Db 22 AAAGTCTTTGAAAGCTTTA 4
|||||

RESULT 5
AI399641/c 697 bp mRNA EST 30-MAR-1999
LOCUS th28a09.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2119576 3',
DEFINITION similar to TR:Q13545 Q13545 SIGNALING INOSITOL POLYPHOSPHATE 5
PHOSPHATASE SIP-145 ;, mRNA sequence.
ACCESSION AI399641
VERSION AI399641.1 GI:4242728
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 697)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2979819.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 832 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 339.
Location/Qualifiers
1. 697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2119576"
/clone_lib="NCI_CGAP_Pr28"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss

```





<hr/>					
AA243694/C	421 bp	mRNA	EST	07-MAR-1997	
LOCUS	zr6e08.s1 Soares_NHMPu_s1 Homo sapiens cDNA clone IMAGE:668582				
DEFINITION	3', mRNA sequence.				
ACCESSION	AA243694				
VERSION	AA243694.1	GI:1874486			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 421)				
AUTHORS	Hillier,L., Clark,N., Dubuca,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfig,T., Soares,M., Tan,F., Trevasik,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.				
TITLE	The WashU-Werck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	On May 9, 1995 this sequence version replaced gi:802693. Contact: Wilton RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL : contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -4lnl3 fwd. ET from Amersham High quality sequence stop: 373.				
FEATURES	Location/Qualifiers				
	1..421				
	/organism="Homo sapiens"				
	/db_xref="CDB:5562549"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:668582"				
	/clone.lib="Soares_NHMPu_s1"				
	/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"				
	/lab_host="DH10B"				
	/note="Organ: liver (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBHL19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."				
	152 a 79 c 55 g 135 t				
BASE COUNT					
ORIGIN					
	Query Match	16.7%; Score 381; DB 30; Length 421;			
	Best Local Similarity	100.0%; Pred. No. 5.7e-155;			
	Matches 381; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1868	ATTCTCCATGTA CTGGCAAGACC TGATTTC TGACCATTTA ATGTGATCG CGTGGA GTA 1927			
Db	383	ATTCTCCATGTA CTGGCAAGACC TGATTTC TGACCATTTA ATGTGATCG CGTGGA GTA 324			
QY	1928	CAAAAGTGGAGT GTGCCCTG AGTAATGC ATTATGGGTGG TTTACCAT TTCTTGAGGT AAA 1987			
Db	323	CAAAAGTGGAGT GTGCCCTG AGTAATGC ATTATGGGTGG TTTACCAT TTCTTGAGGT AAA 264			
QY	1988	AGCATCACAT GAAC TTGTAAGGA AATTAAAA RCCTAC TTTCATA ATAAGTTGC ATAGG 2047			
Db	263	AGCATCACAT GAAC TTGTAAGGA AATTAAAA RCCTAC TTTCATA ATAAGTTGC ATAGG 204			
QY	2048	TTTTAATAAT TTTTAAT ATATATG GCCTTGAG TTGATTTA AATTGT AAATAGG CGGTAA CTAATTTT AA 2107			

Db	300	TAATGCATATAGGGTGGTTTACCAATTCTCTGAGGTAAAGGATACACATGAACCTTGTAAG	241
QY	2010	GAATTAAAAATCCCTACTCTTCATATAAGTTGCATAGGTTTAAATAATTTTAAATATATG	2069
Db	240	GAATTTAAAAATCCCTACTCTTCATATAAGTTGCATAGGTTTAAATAATTTTAAATATATG	181
QY	2070	GCTTGAGTTTAAATGTAATAGGCGTAACTAATTTTAACTCTATATAGTGTTCAATCTGG	2129
Db	180	GCTTGAGTTTAAATGTAATAGGCGTAACTAATTTTAACTCTATATAGTGTTCAATCTGG	121
QY	2130	ATAATCTCTAAACATATGAATTTATGTTTGATCTTTCACCTTCCAAGAGCCCTTTTTTTGAAA	2189
Db	120	ATAATCTCTAAACATATGAATTTATGTTTGATCTTTCACCTTCCAAGAGCCCTTTTTTTGAAA	61
QY	2190	AAAAGCTTTTTTGAATCATCAAGCTTTTCACATTTAAATAAAGCTTTTGAAGCTTTA	2248
Db	60	AAAAGCTTTTTTGAATCATCAAGCTTTTCACATTTAAATAAAGCTTTTGAAGCTTTA	2
RESULT 9			
A0784572/c			
LOCUS			
DEFINITION			
hs_3248_A2_D01.T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3248 Col-2 Row-G, genomic survey sequence.			
ACCESSION			
A0784572			
VERSION			
A0784572.1			
KEYWORDS			
GSS.			
SOURCE			
ORGANISM			
Homo sapiens			
REFERENCE			
AUTHORS			
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.			
TITLE			
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome			
JOURNAL			
MEDLINE			
COMMENT			
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)			
Contact: Mahairas GG, Wallace JC, Hood L			
High Throughput Sequencing Center			
University of Washington			
401 Queen Anne Avenue North, Seattle, WA 98109, USA			
Tel: (206) 616-3618			
Fax: (206) 616-3887			
Email: jwallace@u.washington.edu			
Clones may be purchased from Research Genetics (info@resgen.com).			
BAC end Web Server: http://www.htsc.washington.edu			
Plate: 3248 row: G column: 2			
Seq primer: 77			
Class: BAC ends			
High quality sequence stop: 535.			
FEATURES			
source			
1..535			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="Plate-3248 Col-2 Row-G"			
/clone_lib="CIT Approved Human Genomic Sperm Library D"			
/sex="male"			
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones In E-Coli DH10B"			
BASE COUNT			
ORIGIN			
180 a 90 c 89 g 172 t 4 others			
Query Match			
14.7%; Score 334; DB 87; Length 535;			
Best Local Similarity			
99.7%; Pred. No. 1.3e-134;			
Matches 384; Conservative			
0; Mismatches 1; Indels 0; Gaps 0;			
QY	966	AGCTCTACTGTGTGCTAAAAACAGCAGCTAAACAAAGAGTGAAGGATTTATGTTAAT	1025
Db	535	AGCTGTACTGTGTGCTAAAAACAGCAGCTAAACAAAGAGTGAAGGATTTATGTTAAT	476

the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBHL19W."

BASE COUNT 103 a 108 c 93 g 132 t 1 others

ORIGIN

Query Match 13.8%; Score 314; DB 25; Length 437;  
Best Local Similarity 99.8%; Pred. No. 6.4e-126;  
Matches 434; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 95 ACCAGGCGATGGAGCGAGTGGCTGTATCATGGCGAAATCAGCAGGGAAACCGGGGAGA 154  
|||||  
Db 435 ACCAGGCGATGGAGCGAGTGGCTGTATCATGGCGAAATCAGCAGGGAAACCGGGGAGA 376  
|||||

Qy 155 AGCTCCCTGCTGGCCACTGGGCTGGATGGCAGCTATTGCTGAGGGACAGCGAGCGTGC 214  
|||||

Db 375 AGCTCCCTGCTGGCCACTGGGCTGGATGGCAGCTATTGCTGAGGGACAGCGAGCGTGC 316  
|||||

Qy 215 CAGCGGTGTACTGCGTATGCTGTATCAGCGTTACATTTATACATACCGAGTGTCC 274  
|||||

Db 315 CAGCGGTGTACTGCGTATGCTGTATCAGCGTTACATTTATACATACCGAGTGTCC 256  
|||||

Qy 275 AGACAGAAACAGGTTCTTGAGGTGCTGAGACAGCACCTGGGGTACATAAAGATATTTCC 334  
|||||

Db 255 AGACAGAAACAGGTTCTTGAGGTGCTGAGACAGCACCTGGGGTACATAAAGATATTTCC 196  
|||||

Qy 335 GGAAATAAATAATCTCATTTTC-AGCATTTTCAAGACAGATCAAGGCATTGTATACCT 393  
|||||

Db 195 GGAAATAAATAATCTCATTTTCGAGCATTTTCAAGACAGATCAAGGCATTGTATACCT 136  
|||||

Qy 394 CTGCAGTATCAGTTGAGAGAGTCTCTCAGCTAGAGTACACAGGTACTACAGGGATA 453  
|||||

Db 135 CTGCAGTATCAGTTGAGAGAGTCTCTCAGCTAGAGTACACAGGTACTACAGGGATA 76  
|||||

Qy 454 AGAGAAGATCCTGATGCTCTGAGAGCCCATGAAGCCCATGAAGAAAATAAACAACCTTGTACT 513  
|||||

Db 75 AGAGAAGATCCTGATGCTCTGAGAGCCCATGAAGCCCATGAAGAAAATAAACAACCTTGTACT 16  
|||||

Qy 514 TATTTTCTATAATTT 528  
|||||

Db 15 TATTTTCTATAATTT 1  
|||||

RESULT 11  
AI203763/c 243 bp mRNA EST 27-JAN-1999  
LOCUS qf76e08.x1 Soares\_fetal\_lung\_NbHL19W Homo sapiens CDNA clone  
DEFINITION IMAGE:1755974 3', mRNA sequence.  
ACCESSION AI203763  
VERSION AI203763.1 GI:3756369  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 243)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152111.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1066 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 225.

FEATURES  
Source Location/Qualifiers  
1. .243  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1755974"  
/clone.lib="Soares\_fetal\_lung\_NbHL19W"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: lung; Vector: pT73D (Pharmacia) with a modified polylinker; Site: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I, and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBHL19W." 74 t  
BASE COUNT 88 a 38 c 43 g 132 t  
ORIGIN

Query Match 10.6%; Score 242; DB 43; Length 243;  
Best Local Similarity 100.0%; Pred. No. 1.4e-94;  
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1109 TAAATATGCTTAACATGTTTATCGTATTGATGCTACAGGATTTGAAATTCGTATTAC 1168  
|||||

Db 243 TAAATATGCTTAACATGTTTATCGTATTGATGCTACAGGATTTGAAATTCGTATTAC 184  
|||||

Qy 1169 AAATCCAATGAATGAGTTTCTTTTTCATTTACCTCTGCCCGAGTTGTTCTACTACAT 1228  
|||||

Db 183 AAATCCAATGAATGAGTTTCTTTTTCATTTACCTCTGCCCGAGTTGTTCTACTACAT 124  
|||||

Qy 1229 GGAAGACCTCATTTTGAAGGGAATTTTCAGAGCTGCGAGCTCATGAGTAACTGATTGTA 1288  
|||||

Db 123 GGAAGACCTCATTTTGAAGGGAATTTTCAGAGCTGCGAGCTCATGAGTAACTGATTGTA 64  
|||||

Qy 1289 ACAAGGCTCCTTTTAAAGTAAACCTACAAAACCACTGGAAGTTTATGTTGTTATTATT 1348  
|||||

Db 63 ACAAGGCTCCTTTTAAAGTAAACCTACAAAACCACTGGAAGTTTATGTTGTTATTATT 4  
|||||

Qy 1349 TT 1350  
||

Db 3 TT 2  
||

RESULT 12  
AA233116 208 bp mRNA EST 28-FEB-1997  
LOCUS Zr68e08.r1 Soares\_NHMPu\_S1 Homo sapiens CDNA clone IMAGE:668582  
DEFINITION 5', mRNA sequence.  
ACCESSION AA233116  
VERSION AA233116.1 GI:1856110  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 208)  
AUTHORS Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Dec 3, 1996 this sequence version replaced gi:1122614.  
Contact: Wilton RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810

Email: estewatson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28ml3 rev2 Et from Amersham  
High quality sequence stop: 192.

#### FEATURES

source

Location/Qualifiers  
1...208  
/organism="Homo sapiens"  
/db\_xref="GDB:562549"  
/db\_xref="taxon:9606"  
/clone="IMAGE:668582"  
/clone\_lib="Soares\_Nhrmpu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); vector: pT7T3D-Pac  
(pharmacia) with a modified polylinker; Site\_1: Not I;  
Site\_2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBHM, pregnant uterus  
NBHPU, and fetal heart NBH19W) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of I.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."  
73 a 35 c 38 g 62 t

#### BASE COUNT

ORIGIN

Query Match 7.1%; Score 162; DB 30; Length 208;  
Best Local Similarity 100.0%; Pred. No. 7.4e-60;  
Matches 162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 464 CTGATGTCGCCGGAAGCCCATGAAGAAATAAACAACCTGTACTTTTCTAT 523  
|||||  
Db 47 CTGATGTCGCCGGAAGCCCATGAAGAAATAAACAACCTGTACTTTTCTAT 106  
|||||  
Qy 524 AATTAAATATATGCTAAGCTTATATTTAGATATACAGTTCGGTGAGCTACAAAT 583  
|||||  
Db 107 AATTAAATATATGCTAAGCTTATATTTAGATATACAGTTCGGTGAGCTACAAAT 166  
|||||  
Qy 584 GCATTCTAAGCCATGTAGTCTGTATGATGAGCATCTAG 625  
|||||  
Db 167 GCATTCTAAGCCATGTAGTCTGTATGATGAGCATCTAG 208  
|||||

#### RESULT 13

AA190204/c

LOCUS

DEFINITION

TH075 HPCDL1 Homo sapiens cDNA 5'/3' similar to Unknown, mRNA

sequence.

ACCESSION

AA190204

VERSION

AA190204.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 166)

Authors

Sohn, U., Park, D.S., Lee, C.M., Cho, W.K., Ahn, H.J., Lee, M.Y.,

Hwang, M.Y. and Jin, S.W.

Title

Human HPCDL1 library cDNAs

Journal

Unpublished (1994)

Comment

On Sep 12, 1996 this sequence version replaced gi:1407222.

Contact: Uik

Sohn, D.S. Park, C.M. Lee, W. K. Cho, H. J. Ahn, M. Y. Lee, M. Y. Hwang, S. W. Jin

Laboratory of Molecular Biology

Kyungpook National University

Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701,

Korea

Tel: 82-053-950-5382  
Fax: 82-053-955-5327

Email: usohn@bh.kyungpook.ac.kr  
Seq primer: M13 Reverse/SK primer.

#### FEATURES

source

Location/Qualifiers  
1...166  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HTCBL1"  
/lab\_host="XLI-Blue"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
EcoRI; Poly(A)-mRNA from the 2-year old male fetal  
thymus, oligo(dT) priming, EcoRI cloning in the vector  
pBluescript (Stratagene)."  
66 a 18 c 24 g 58 t

#### BASE COUNT

ORIGIN

Query Match 6.8%; Score 155; DB 29; Length 166;  
Best Local Similarity 100.0%; Pred. No. 8.8e-57;  
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2098 CTAATTTAACTCTATAATGTGTCATTCGGAATATCCTAAACATATGAATATGTTT 2157  
|||||  
Db 166 CTAATTTAACTCTATAATGTGTCATTCGGAATATCCTAAACATATGAATATGTTT 107  
|||||  
Qy 2158 GCATGTCACCTCCAGAGCCCTTTTGGAAAAAGCTTTTGAATCAATCAAGTCTT 2217  
|||||  
Db 106 GCATGTCACCTCCAGAGCCCTTTTGGAAAAAGCTTTTGAATCAATCAAGTCTT 47  
|||||

Qy 2218 TCACATTTAAATAAGTGTGTTGAAAGCTTTATTTA 2252  
|||||

Db 46 TCACATTTAAATAAGTGTGTTGAAAGCTTTATTTA 12  
|||||

#### RESULT 14

AW001042/c

LOCUS

DEFINITION

wr91h07.xl NCI-CGAP Kid11 Homo sapiens cDNA clone IMAGE:2495101 3',

mRNA sequence.

ACCESSION

AW001042

VERSION

AW001042.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 359)

Authors

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Title

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Journal

Unpublished (1997)

Comment

On May 9, 1996 this sequence version replaced gi:1132712.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.

Location/Qualifiers

1...359

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2495101"

/clone\_lib="NCI-CGAP\_Kid11"

/lab\_host="DH10B"

/note-Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500352-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 120 a 60 c 60 g 119 t

ORIGIN

Query Match 5.4%; Score 122; DB 63; Length 359;  
Best Local Similarity 100.0%; Pred. No. 1.3e-42; Indels 0; Gaps 0;  
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 570 GGTGAGCTCAATGATTTCTAAGCCATTGTAGTCTGTATGGAAGCATCTAGCATG 629  
Db 246 GGTGAGCTCAATGATTTCTAAGCCATTGTAGTCTGTATGGAAGCATCTAGCATG 187  
Qy 630 TCGTCAAGCTGAATGGACTTTTGTACATAGTGTAGGAGCTTTGAAACGAGGATTGGAA 689  
Db 186 TCGTCAAGCTGAATGGACTTTTGTACATAGTGTAGGAGCTTTGAAACGAGGATTGGAA 127  
Qy 690 AA 691  
Db 126 AA 125

RESULT 15  
AA585096  
LOCUS 172 bp mRNA EST 09-SEP-1997  
DEFINITION ATH247 HTCDLL Homo sapiens CDNA 5'/3', mRNA sequence.  
ACCESSION AA585096  
VERSION AA585096.1 GI:2384984  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 172)  
AUTHORS Sohn, U., Park, D.S., Lee, C.M., Cho, W.K., Ahn, H.J., Lee, M.Y., Hwang, M.Y. and Jin, S.W.  
TITLE Human HTCDLL library cDNAs  
JOURNAL Unpublished (1994)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1294001.  
Contact: Uik  
Sohn, D.S. Park, C.M. Lee, W.K. Cho, H.J. Ahn, M.Y. Lee, M.Y. Hwang, S.W. Jin  
Laboratory of Molecular Biology  
Kyungpook National University  
Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701, Korea  
Tel: 82-053-950-5382  
Fax: 82-053-955-5327  
Email: usongm@kyungpook.ac.kr  
Seq primer: M13 Reverse/SK primer.  
Location/Qualifiers  
1. 172  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HTCDLL"  
/lab\_host="Xtl-Blue"  
/note="vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2: EcoRI; Poly(A)-mRNA from the 2-year old male fetal thymus, oligo(dT) priming, EcoRI cloning in the vector pBluescript (Stratagene)."

BASE COUNT 54 a 27 c 27 g 64 t

ORIGIN

Query Match 5.2%; Score 119; DB 35; Length 172;  
Best Local Similarity 100.0%; Pred. No. 8e-38;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2099 TAATTTAACTCATATGTTCTGTTCTGGAATATCTAAACATATGAATTTGTTG 2158  
Db 29 TAATTTAACTCATATGTTCTGTTCTGGAATATCTAAACATATGAATTTGTTG 88  
Qy 2159 CAGTTCACCTCCAGAGCCTTTTTTGAAGAAAAGCTTTTTTGAATCATCAAGTCTT 2217  
Db 89 CAGTTCACCTCCAGAGCCTTTTTTGAAGAAAAGCTTTTTTGAATCATCAAGTCTT 147

RESULT 16  
AA354319  
LOCUS 331 bp mRNA EST 21-APR-1997  
DEFINITION Jurkat T-cells V Homo sapiens CDNA 5' end, mRNA sequence.  
ACCESSION AA354319  
VERSION AA354319.1 GI:2006639  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 331)  
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Well, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bedarik, D.P., Caci, L., Cepeda, M.A., Coleman, T.A., Collins, F.J., Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.  
TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)  
MEDLINE 12140200  
COMMENT On May 8, 1995 this sequence version replaced gi:801515.  
Other\_ESRs: THCL48776  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression information related to this ESR, please check the TIGR Human Gene Index (<http://www.tigr.org/tldb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 331  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):154561"  
/db\_xref="taxon:9606"  
/clone\_lib="Jurkat T-cells v"  
/cell\_type="T-lymphocyte"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 84 a 82 c 89 g 73 t 3 others

ORIGIN

Query Match 4.9%; Score 111; DB 32; Length 331;  
Best Local Similarity 100.0%; Pred. No. 8e-38;  
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

102 CATGACGACGTGCTGTGTCATCATGCAAAATCAGCAGGAAACCGGAGAGCTCCT 161  
QY  
49 CATGACGACGTGCTGTGTCATCATGCAAAATCAGCAGGAAACCGGAGAGCTCCT 108  
Db  
162 GCTTCCCACTGGCTGGATGCACGCTATTTGCTGAGGCACAGCAGAGCT 212  
QY  
109 GCTTCCCACTGGCTGGATGCACGCTATTTGCTGAGGCACAGCAGAGCT 159  
Db

RESULT	17	
LOCUS	T25937	
DEFINITION	T25937 150 bp mRNA EST 10-MAR-1998 ATRH37 HCTDL1 Homo sapiens cDNA 5'/3' similar to ATRase 1, mRNA sequence.	
ACCESSION	T25937	
VERSION	T25937.1	GI:2947531
KEYWORDS	EST.	
SOURCE	human.	

SOURCE	ORGANISM	Homosapiens
Human.	Human.	Human.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

**REFERENCE**

Sachdev, R. N. S., 1986, *Physica Scripta*, T17, 1.

**AUTHORS**

Sohn, U., Park, D. S., Lee, C. M., Cho, W. K., Ahn, H. J., Lee, M. Y.,  
1 (bases 1 to 150)

**TITLE**  
Hwang, M.Y. and Jin, S.W.  
Human HTCDL1 library cDNAs  
Unpublished (1994)

COMMENT

Contact: Uik  
Sohn, D.S. Park, C.M. Lee, W.K. Cho, H.J. Ahn, M.Y. Lee, M.Y. Hwang, S.W. Jin  
Laboratory of Molecular Biology  
Kyungpook National University  
Dept. of Genetic Eng., Kyungpook National Univ., Taegu 702-701,  
Korea  
Tel: 82-053-950-5382  
Fax: 82-053-955-5327  
Email: usohn@kh.kyungpook.ac.kr  
EST is putatively homologous to ATPase 1 (PIR, A44396) in  
Plasmodium falciparum with 50% (5 amino acids overlapped) identity.  
Seq primer: M13 Reverse/SK primer.

RESULT	18
W19453	
LOCUS	W19453
DEFINITION	477 bp mRNA EST
IMAGE:	302800 5', mRNA sequence.
ACCESSION	W19453
	ZB22C09.r1 Soares_fetal_lung_NHHL19W Homo sapiens cDNA clone
	03-MAY-1996

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

W19453.1 GI:1295392  
EST.  
human.  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 477)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlffing, T., Soares, M., Tan, F.,  
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE The Wabhu-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On Jan 24 1995 this sequence version replaced gl:634373.  
Contact: Wilson FK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

```

tax: 001000
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 216.

Location/Qualifiers
1..477
/organism="Homo sapiens"
/db_xref="GDB:1247732"
/db_xref="taxon:9606"
/clone="IMAGE:302800"
/clone_lib="Soares.fetal_lung_NbHL19w"
/dev_stage="19 weeks"
/lab_host="Df10B (ampicillin resistant)"
/notes="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer
[5'-TGTTCAATCTGAATGGGACGCGCCATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fátima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19w"

```

BASE COUNT	138 a	108 c	114 g	110 t	7 others
ORIGIN					
Query Match		3.2%	Score 73;	DB 26;	Length 477;
Best Local Similarity		99.2%;	Prod. No. 2.1e-21;		
Matches 123;	Conservative		0: Mismatches	1: Indels	0: Gaps

RESULT	19				
LOCUS	AA355783				
DEFINITION	AA355783	308 bp	mRNA	EST	21-Apr-1997
ACCESSION	AA355783	EST64284	Jurkat T-cells	VI Homo sapiens	CDNA 5' end, mRNA sequence.
VERSION	AA355783.1	GI:2008144			

```

KEYWORDS  EST.
SOURCE      human.
ORGANISM    Homo sapiens

REFERENCE   1 (bases 1 to 308)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
             Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
             White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whai,C.,
             Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
             Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
             Glodke,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
             Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
             Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
             Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
             Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
             Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
             Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
             He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
             Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
             Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
             Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
             Fraser,C.M. and Venter,J.C.
TITLE       Initial assessment of human gene diversity and expression patterns
             based upon 83 million nucleotides of cDNA sequence
JOURNAL     Nature 377 (6547 suppl), 3-174 (1995)
MEDLINE     12140200
COMMENT     On Apr 14, 1993 this sequence version replaced gi:638038.
             Other_ESTs: THCL48776
             Contact: Kerlavage, AR
             Bioinformatics
             The Institute for Genomic Research
             9712 Medical Center Drive, Rockville, MD 20850 USA
             Tel: 3018699056
             Fax: 3018699423
             Email: arkerlav@tigr.org
             For clone availability, additional sequence and expression
             information related to this EST, please check the TIGR Human Gene
             Index (http://www.tigr.org/tdb/hgi/hgi.html)
             Seq primer: M13 Reverse.

FEATURES    source
             Location/Qualifiers
             1..308
             /organism="Homo sapiens"
             /db_xref="ATCC (inhost):157953"
             /db_xref="taxon:9606"
             /clone_lib="Jurkat T-cells v1"
             /cell_type="T-lymphocyte"
             /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
             XhoI"
BASE COUNT  69 a 70 c 87 g 76 t 6 others
ORIGIN
Query Match 2.0%; Score 46; DB 32; Length 308;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGAACGGGAGTCAGCTGCTGACTTGTCCTGGCTGCAGTAGCA 46
Db 146 CTGAACGGGAGTCAGCTGCTGACTTGTCCTGGCTGCAGTAGCA 191
|||||
|||||

RESULT 20
LOCUS     AI030232/c
DEFINITION UI-R-CO-it-a-04-0-UI sl UI-R-CO Rattus norvegicus cDNA clone
DEFINITION UI-R-CO-it-a-04-0-UI 3', mRNA sequence.
ACCESSION AI030232
VERSION   AI030232.1 GI:3248058
KEYWORDS  EST.
SOURCE    Norway rat.
ORGANISM  Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE   1 (bases 1 to 590)
AUTHORS     Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
TITLE       Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     On Jan 17, 1998 this sequence version replaced gi:1900524.
             Program for Rat Gene Discovery and Mapping
             University of Iowa
             451 Eckstein Medical Research Building Iowa City, IA 52242, USA
             Tel: 319 335 8250
             Fax: 319 335 9565
             Email: msoares@blue.weeg.uiowa.edu
             The sequence tag present in the cDNA between the NotI site and the
             oligo-dT track served to identify it as a clone from the normalized
             adult 8-Day-Embryo library. cDNA Library Preparation: M. Fatima
             Bonaldo, Ph.D. Clone distribution: clones will be available through
             Research Genetics
             Seq primer: M13 Forward.
             Location/Qualifiers
             1..590
             /organism="Rattus norvegicus"
             /strain="Sprague-Dawley"
             /db_xref="taxon:10116"
             /clone="UI-R-CO-it-a-04-0-UI"
             /clone_lib="UI-R-CO"
             /dev_stage="adult"
             /lab_host="DH10B (Life Technologies)"
             /note="Vector: pT7T3D-pac (Pharmacia) with a modified
             polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CO
             library is a subtracted library derived from the UI-R-Al
             and UI-R-El libraries. The UI-R-Al library consisted of a
             mixture of individually tagged normalized libraries
             constructed from rat placenta, adult lung, brain, liver,
             kidney, heart, spleen, ovary, and muscle. The UI-R-El
             library consisted of a mixture of individually tagged
             normalized libraries constructed from 8, 12 and 18-day
             embryo. The tag is a string of 3-5 nucleotides present
             between the Not I site and the oligo-dT track which
             allows identification of the library of origin of a clone
             within the mixture. The subtracted library (UI-R-CO) was
             constructed as follows: PCR amplified cDNA inserts from a
             pool of UI-R-Al and UI-R-El clones from which 3' ESTs had
             been derived was used as a driver in a hybridization with
             the pooled UI-R-Al and UI-R-El library in the form of
             single-stranded circles. The remaining single-stranded
             circles (subtracted library) was purified by
             hydroxyapatite column chromatography, converted to
             double-stranded circles and electroporated into DH10B
             bacteria (Life Technologies) to generate the UI-R-CO
             library. This procedure has been previously described
             (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
             1996)."
BASE COUNT  163 a 116 c 109 g 202 t
ORIGIN
Query Match 1.9%; Score 43; DB 41; Length 590;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 341 TAAARAATCTCATTTTCAGCATTTTCAGAGCCAGATCAAGGCAT 383
Db 488 TAAARAATCTCATTTTCAGCATTTTCAGAGCCAGATCAAGGCAT 446
|||||
|||||

RESULT 21
LOCUS     AI131797
DEFINITION uc37e03.r1 Soares mouse mammary gland NbMMG Mus musculus cDNA clone
DEFINITION IMAGE:1400188 5' similar to TR:Q13545 013545 SIGNALING INOSITOL
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```



POLYPHOSPHATE 5 PHOSPHATASE SIP-145 ;, mRNA sequence.

ACCESSION A1131797 GI:3601813

VERSION A1131797.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS 1 (bases 1 to 271)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1407241.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:911304

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 89.

Location/Qualifiers

1. .271

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:1400188"

/clone\_lib="Soares mouse mammary gland NbMMG"

/sex="male"

/tissue\_type="mammary gland"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: mammary gland; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru KO, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 66 a 63 c 75 g 67 t

ORIGIN

Query Match 1.5%; Score 35; DB 42; Length 271;

Best Local Similarity 100.0%; Pred. No. 8e-05;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 CGAGTGTCACAGACAGAAACAGGTTCTTGGAGTGC 299

|||||

Db 203 CGAGTGTCACAGACAGAAACAGGTTCTTGGAGTGC 237

|||||

RESULT 22

AA255258 294 bp mRNA EST 14-MAR-1997

LOCUS mz77h04.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone

DEFINITION IMAGE:719479 5' similar to TR:G1255352 G1255352 SH2-CONTAINING INOSITOL PHOSPHATASE SHIP. ;, mRNA sequence.

ACCESSION AA255258

VERSION AA255258.1 GI:1889847

KEYWORDS EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 294)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Apr 14, 1993 this sequence version replaced gi:638194.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:444975

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 275.

Location/Qualifiers

1. .294

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:719479"

/clone\_lib="Soares mouse lymph node NbMLN"

/sex="male"

/tissue\_type="lymph node"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: lymph node; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 80 a 69 c 76 g 69 t

ORIGIN

Query Match 1.5%; Score 35; DB 30; Length 294;

Best Local Similarity 100.0%; Pred. No. 7.8e-05;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 CGAGTGTCACAGACAGAAACAGGTTCTTGGAGTGC 299

|||||

Db 84 CGAGTGTCACAGACAGAAACAGGTTCTTGGAGTGC 118

|||||

RESULT 23

AI464772 397 bp mRNA EST 09-MAR-1999

LOCUS mz77h04.y1 Soares mouse lymph node NbMLN Mus musculus cDNA clone

DEFINITION IMAGE:719479 5' similar to TR:Q61181 Q61181 INOSITOL POLYPHOSPHATE-5-PHOSPHATASE, 145 KDA ;, mRNA sequence.

ACCESSION AI464772

VERSION AI464772.1 GI:4318802

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 397)

## AUTHORS

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

On May 7, 1998 this sequence version replaced gi:3121624.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

This read is a RESEQUENCE of a previously sequenced mouse clone

This read has been verified (found to hit its original self in the

correct orientation)

Putative full length read

vector to vector length is 643

Seq primer: -40RP from Gibco

High quality sequence stop: 362.

Location/Qualifiers

1. .397

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:719479"

/clone\_lib="Soares mouse lymph node NbMLN"

/sex="male"

/tissue\_type="lymph node"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer

[5,

TGTTACCAATCTGAAGTGGAGCGCGCGATACACTTTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. RNA

provided by Dr. Bertrand Jordan. Library constructed and

normalized by Bento Soares and M.Fatima Bonaldo."

113 a 87 c 95 g 102 t

## BASE COUNT

## ORIGIN

Query Match 1.5%; Score 35; DB 47; Length 397;

Best Local Similarity 100.0%; Pred. No. 6.9e-05; Indels 0; Gaps 0;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 265 CGAGTGTCACAGACAGAACACAGGTCCTTGAGTGC 299

Db 84 CGAGTGTCACAGACAGAACACAGGTCCTTGAGTGC 118

## RESULT 24

## AI522369/c

## LOCUS

fb22ell.x1 zebrafish WashU MPIMG EST Danio rerio cDNA 3', mRNA

sequence. 18-MAR-1999

AI522369.1

GI:4436504

EST.

zebrafish.

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;

Cyprinidae; Cyprinidae; Rasbora; Danio.

1 (bases 1 to 210)

Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,

Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

WashU zebrafish EST Project 1998

Unpublished (1998)

On Mar 10, 1998 this sequence version replaced gi:2949061.

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7 ET from Amersham

High quality sequence stop: 187.

Location/Qualifiers

1. .210

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone\_lib="zebrafish WashU MPIMG EST"

/sex="mixed"

/tissue\_type="26 somite embryos, adult livers, shield

stage embryos"

/lab\_host="XLI-blue MRF"

/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; 1st

strand cDNA was primed with a Not I - oligo(dT)15 primer

15'GCAGTACTCTAGATCGGAGCGCGCCCTTTTTTTTTTTT3'];

double-stranded cDNA was ligated to Sal I adaptors (BRL),

digested with Not I and cloned into the Not I and Sal I

sites of the pSPORT1 vector (BRL). Library was constructed

by Matthew Clark (Lehrach lab; ICRF, London and Max Planck

Institut fuer Molekulare Genetik, Berlin). cDNAs for EST

analysis were selected following oligonucleotide

hybridization fingerprinting of arrayed clones from

zebrafish late somitogenesis (26 ss) adult liver or

embryonic shield stage (5.6 h) libraries. Fingerprint

data were used to computationally cluster cDNAs, and a

single cDNA from each cluster was chosen for sequencing.

In some cases multiple members of the same cluster were

sequenced to assess clustering parameters or single clones

were sequenced additional times to assess quality

control."

BASE COUNT 72 a 28 c 25 g 85 t

ORIGIN

Query Match 1.4%; Score 33; DB 47; Length 210;

Best Local Similarity 100.0%; Pred. No. 0.00066;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2244 CTTTATTAAAAAATAAAAAAAAAAAAAAAAAA 2276

Db 33 CTTTATTAAAAAATAAAAAAAAAAAAAAAAAA 1

## RESULT 25

## AU074326

## LOCUS

DEFINITION

AU074326 Dictyostellium discoidium SS (H.Urushihara) Dictyostellium

discoidium cDNA clone SSK377, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

237 bp mRNA EST 24-JUN-1999

237 bp mRNA EST 24-JUN-1999



(web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
 RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))

Possible reversed clone: polyt not found

Seq primer: T7 ET from Amersham

High quality sequence stop: 408.

#### FEATURES

```

source
  1. 420
  /location/Qualifiers
  /organism="Danio rerio"
  /db_xref="taxon:7955"
  /clone_lib="zebrafish washu MPIMG EST"
  /sex="mixed"
  /tissue_type="26 somite embryos, adult livers, shield
  stage embryos"
  /lab_host="XLI-blue MRF"
  /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
  strand cDNA was primed with a Not I - oligo(dT)15 primer
  [5'-pgactagttctagatcgacgagcgccgctttttttttttt3'];
  double-stranded cDNA was ligated to Sal I adaptors (BRL),
  digested with Not I and cloned into the Not I and Sal I
  sites of the pSPORT1 vector (BRL). Library was constructed
  by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
  Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
  analysis were selected following oligonucleotide
  hybridization fingerprinting of arrayed clones from
  zebrafish late somitogenesis (26 ss), adult liver or
  embryonic shield stage (5.6 h) libraries. Fingerprint
  data were used to computationally cluster cDNAs, and a
  single cDNA from each cluster was chosen for sequencing.
  In some cases multiple members of the same cluster were
  sequenced to assess clustering parameters or single clones
  were sequenced additional times to assess quality
  control."
```

BASE COUNT 108 a 91 c 67 g 151 t 3 others

Query Match 1.4%; Score 33; DB 48; Length 420;  
 Best Local Similarity 100.0%; Pred. NO. 0.00049;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276  
 |||||||||||||||||||||||||||||||||||||||  
 Db 45 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 13

RESULT 28  
 AI721672/c  
 LOCUS fc30g11.x1 zebrafish Washu MPIMG EST Danio rerio cDNA 3', mRNA  
 DEFINITION  
 ACCSSION  
 VERSION fc30g11.1 GI:5040001  
 KEYWORDS  
 SOURCE EST.  
 ORGANISM zebrafish.  
 Danio rerio  
 Eukaryota; Chordata; Craniata; Vertebrata; Actinopterygii;  
 Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
 Cyprinidae; Cyprininae; Rasbora; Danio.  
 1 (bases 1 to 424)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 Washu Zebrafish EST Project 1998  
 Unpublished (1998)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 On Mar 10, 1998 this sequence version replaced gi:2947976.  
 Other ESTs: fc30g11.y1  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [zbrafish@watson.wustl.edu](mailto:zbrafish@watson.wustl.edu)  
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,  
 Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
 (web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
 RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))

Seq primer: T7 ET from Amersham  
 High quality sequence stop: 423.

#### FEATURES

```

source
  1. 424
  /location/Qualifiers
  /organism="Danio rerio"
  /db_xref="taxon:7955"
  /clone_lib="zebrafish Washu MPIMG EST"
  /sex="mixed"
  /tissue_type="26 somite embryos, adult livers, shield
  stage embryos"
  /lab_host="XLI-blue MRF"
  /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
  strand cDNA was primed with a Not I - oligo(dT)15 primer
  [5'-pgactagttctagatcgacgagcgccgctttttttttttt3'];
  double-stranded cDNA was ligated to Sal I adaptors (BRL),
  digested with Not I and cloned into the Not I and Sal I
  sites of the pSPORT1 vector (BRL). Library was constructed
  by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
  Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
  analysis were selected following oligonucleotide
  hybridization fingerprinting of arrayed clones from
  zebrafish late somitogenesis (26 ss), adult liver or
  embryonic shield stage (5.6 h) libraries. Fingerprint
  data were used to computationally cluster cDNAs, and a
  single cDNA from each cluster was chosen for sequencing.
  In some cases multiple members of the same cluster were
  sequenced to assess clustering parameters or single clones
  were sequenced additional times to assess quality
  control."
```

BASE COUNT 106 a 94 c 92 g 132 t

Query Match 1.4%; Score 33; DB 51; Length 424;  
 Best Local Similarity 100.0%; Pred. NO. 0.00049;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2244 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276  
 |||||||||||||||||||||||||||||||||||||||  
 Db 44 CTTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 12

RESULT 29  
 AI833180/c  
 LOCUS at65f03.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone  
 DEFINITION IMAGE:2376893 3', similar to gb:X61100 NADH-UBIQUINONE  
 OXIDOREDUCTASE 75 KD SUBUNIT PRECURSOR (HUMAN); mRNA sequence.  
 AI833180  
 VERSION AI833180.1 GI:5455160  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 553)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,  
 Krizman, D., Kucaba, T., Lacy, M., Lennon, G., Marra, M.,  
 Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
 Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
 Washu-NCI human EST Project  
 Unpublished (1997)

## COMMENT

On Dec 20, 1995 this sequence version replaced gi:1134468.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 220.

## FEATURES

## Source

1..553  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2376893"  
 /clone\_lib="Barstead colon HPLRB7"  
 /sex="male"  
 /dev\_stage="adult, age 25"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - Oligo(dT) primer [5'  
 TGTTACGAATCTGAAGTGGAGCGCCCTTTTTTTTTTTTTTTTTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [5' AATTCACCTAGTAAT 3' and 5' ATTACTAGTG 3'], digested  
 with Not I and cloned into the Not I and Eco RI sites of  
 the modified pT73 vector. Library constructed by Bob  
 Barstead."

## BASE COUNT

156 a 121 c 97 g 179 t

## ORIGIN

Query Match 1.4%; Score 33; DB 61; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0.00044; Indels 0; Gaps 0;  
 Matches 33; Conservative 0; Mismatches 0;  
 QY 2244 CTTTATTATAAAAAAAAAAAAAAAAAAAAAA 2276  
 |||||||||||||||||||||||||||||||||||  
 Db 37 CTTTATTATAAAAAAAAAAAAAAAAAAAAAA 5

## RESULT 30

AA503384/C  
 LOCUS  
 DEFINITION ne45b11.sl NCI\_CGAP\_Co3 Homo sapiens cDNA clone IMAGE:900285 3',  
 mRNA sequence.  
 ACCESSION AA503384  
 VERSION AA503384.1 GI:2238351  
 KEYWORDS EST  
 SOURCE human.  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 75)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1995 this sequence version replaced gi:111178.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue procurement: Elias Campo, M.D., Michael R. Emmert-Buck,  
 M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 772 Std Error: 0.00

## FEATURES

## Source

Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 51.  
 Location/Qualifiers  
 1..75  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:900285"  
 /clone\_lib="NCI\_CGAP\_Co3"  
 /sex="pooled"  
 /tissue\_type="colon"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from 12 pooled bulk tumor samples and primed  
 with a Not I - Oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT73 vector. Library went through one round of  
 normalization."

## BASE COUNT

10 a 2 c 1 g 62 t

## ORIGIN

Query Match 1.4%; Score 32; DB 34; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 0.0027;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2245 TTTATTATAAAAAAAAAAAAAAAAAAAAAA 2276  
 |||||||||||||||||||||||||||||||||||  
 Db 70 TTTATTATAAAAAAAAAAAAAAAAAAAAAA 39

## RESULT 31

AI960856  
 LOCUS  
 DEFINITION sc9le07.yl Gm-cl019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl019-733 5', mRNA sequence.  
 ACCESSION AI960856  
 VERSION AI960856.1 GI:5753569  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM  
 Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core  
 eudicotids; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 Glycine.  
 REFERENCE 1 (bases 1 to 160)  
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Corvelli,V.,  
 Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,  
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,  
 Bowers,J., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,  
 Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,N.,  
 McCann,R., Waterston,R. and Wilson,R.  
 TITLE Public Soybean EST Project  
 JOURNAL Unpublished (1999)  
 COMMENT On Mar 16, 1998 this sequence version replaced gi:2961824.  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World  
 Parkway Circle St. Louis, Missouri 63134 For further information  
 call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)  
 427-3324 or contact: clones@genomesystems.com or  
 info@genomesystems.com web site: www.genomesystems.com  
 putative full length read  
 vector to vector length is 804  
 High quality sequence stop: 132.

## FEATURES

## Source

1..160

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-733"
/clone_lib="Gm-cl019"
/tissue_type="immature seed coats of greenhouse grown
plants"
/lab_host="DH10B (Gibco BRL)"
/note="vector: pSPORT1 (Life Technologies); Site_1: Not I;
Site_2: Sal I; This cDNA library was constructed from mRNA
isolated from immature seed coats (200-300 mgs) of
greenhouse grown plants. The library was prepared using
the Life Technologies pSuperscript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a Not I restriction
site. Sal I linker adapters were ligated to the
blunt-ended cDNA fragments followed by Not I digestion.
The cDNA fragments were directionally cloned into the Not
I-Sal I restriction site of the pSPORT1 vector. The
ligated cDNA fragments were transformed into E.coli
Electromax DH10B host cells (Gibco BRL). This library was
constructed by Dr. Lilia Vodkin and Dr. Anu Khanna."
BASE COUNT      100 a      10 c      16 g      34 t
ORIGIN

Query Match      1.4%; Score 32; DB 63; Length 160;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA.....AAAAAAAAAAAAA 2276
|||||.....|
DB 77 TTTATTTAAAAA.....AAAAAAAAAAAAA 108
|||||.....|

RESULT 32
LOCUS AU037642 168 bp mRNA EST 29-MAR-1999
DEFINITION AU037642 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium
discoideum cDNA clone SSE122, mRNA sequence.
ACCESSION AU037642
VERSION AU037642.1 GI:3984395
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE Eukaryota; Dictyostellida; Dictyostellium.
AUTHORS Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochial,H. and Tanaka,Y.
TITLE The Dictyostellium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
JOURNAL DNA Res. 5 (6), 335-340 (1998)
MEDLINE 99156227
COMMENT On Jan 19, 1998 this sequence version replaced gi:2284680.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@akura.cc.tsukuba.ac.jp
PROJECT - 'Dictyostellium discoideum cDNA project in Japan'.
FEATURES
source
1. .168
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSE122"
/clone_lib="Dictyostellium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT      89 a      20 c      6 g      50 t      3 others
ORIGIN

```

```

Query Match      1.4%; Score 32; DB 44; Length 168;
Best Local Similarity 100.0%; Pred. No. 0.002;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA.....AAAAAAAAAAAAA 2276
|||||.....|
DB 131 TTTATTTAAAAA.....AAAAAAAAAAAAA 162
|||||.....|

RESULT 33
LOCUS C94407 174 bp mRNA EST 15-JUN-1998
DEFINITION C94407 Dictyostellium discoideum SS (H.Urushihara) Dictyostellium
discoideum cDNA clone SSK659, mRNA sequence.
ACCESSION C94407
VERSION C94407.1 GI:3219022
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE Eukaryota; Dictyostellida; Dictyostellium.
AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostellium discoideum
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797274.
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@akura.cc.tsukuba.ac.jp
PROJECT - Dictyostellium discoideum cDNA project in Japan.
FEATURES
source
1. .174
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSK659"
/clone_lib="Dictyostellium discoideum SS (H.Urushihara)"
/dev_stage="slug"
BASE COUNT      108 a      12 c      8 g      46 t
ORIGIN

Query Match      1.4%; Score 32; DB 41; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTTAAAAA.....AAAAAAAAAAAAA 2276
|||||.....|
DB 117 TTTATTTAAAAA.....AAAAAAAAAAAAA 148
|||||.....|

RESULT 34
LOCUS AI811150/c 178 bp mRNA EST 07-JUL-1999
DEFINITION AI811150 x1 NCI_CGAP_Utl Homo sapiens cDNA clone IMAGE:2261894 3',
mRNA sequence.
ACCESSION AI811150
VERSION AI811150.1 GI:5397716
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS 1 (bases 1 to 178)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948846.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

```

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco.

FEATURES  
 source  
 1. .178  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2261894"  
 /clone\_lib="NCI-CCAP\_Utl"  
 /tissue\_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"  
 /lab\_host="PH103"  
 /note="Organ: uterus; Vector: PCMV-SPORT6; Site\_1: Sali; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"  
 72 a 24 c 16 g 66 t

Query Match 1.4%; Score 32; DB 61; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276  
 |||  
 Db 32 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 1

RESULT 35  
 AW059084  
 LOCUS  
 DEFINITION  
 fe24h07.x1 Zebrafish Washu MPIMG EST Danio rerio cDNA 3', mRNA  
 sequence.  
 AW059084  
 VERSION  
 AW059084.1 GI:5934723  
 KEYWORDS  
 EST.  
 SOURCE  
 zebrafish.  
 ORGANISM  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
 Cyprinidae; Cyprininae; Rasbora; Danio.  
 1 (bases 1 to 215)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,  
 Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,  
 Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R., and Wilson, R.  
 Washu Zebrafish EST Project 1998  
 Unpublished (1998)  
 On May 18, 1998 this sequence version replaced gi:3138053.  
 Contact: Stephen L. Johnson  
 Washington University School  
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrfish@watson.wustl.edu  
 cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:  
 Matthew Clark. DNA Sequencing by: Washington University Genome  
 Sequencing Center Clone distribution: Genome Systems, St. Louis,  
 Missouri (web address: [www.genomesystems.com](http://www.genomesystems.com)) (email contact:  
[info@genomesystems.com](mailto:info@genomesystems.com)) and Research Genetics, Huntsville, Alabama  
 (web address: [www.resgen.com](http://www.resgen.com)) (email contact: [info@resgen.com](mailto:info@resgen.com)) and  
 RessourcenZentrumPrimaDatenbank, Berlin, Germany (web address:  
[www.rzpd.de](http://www.rzpd.de))

FEATURES  
 source

Seq primer: 77 ET from Amersham.  
 Location/Qualifiers  
 1. .215  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 /clone\_lib="Zebrafish Washu MPIMG EST"  
 /sex="mixed"  
 /tissue\_type="26 somite embryos, adult livers, shield stage embryos"  
 /lab\_host="XLI-blue MRF"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: Sali; 1st strand cDNA was primed with a Not I - oligo(dT)15 primer [5' pGACTAGTCTAGATCCGACGCGCCCTTTTCTTTT3]; double-stranded cDNA was ligated to sal I adaptors (BRL), digested with Not I and cloned into the Not I and sal I sites of the pSPORT1 vector (BRL). Library was constructed by Matthew Clark (Lehrach lab; ICRF, London and Max Planck Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control."  
 104 a 24 c 32 g 55 t

Query Match 1.4%; Score 32; DB 64; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 2276  
 |||  
 Db 157 TTTATTTAAAAA...AAAAAAAAAAAAAAAAAAAA 188

RESULT 36  
 AW059109/C  
 LOCUS  
 DEFINITION  
 wc67e08.x1 NCI-CCAP\_Panl Homo sapiens cDNA clone IMAGE:2323718 3', similar to contains Alu repetitive element; contains element LTR9 repetitive element ;, mRNA sequence.  
 AW059109  
 VERSION  
 AW059109.1 GI:4896403  
 KEYWORDS  
 EST.  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 217)  
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Jun 5, 1998 this sequence version replaced gi:3187518.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: [Robert.strausberg@nih.gov](mailto:Robert.strausberg@nih.gov)  
 Life Technologies catalog #: 11548-013  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 174.  
 Location/Qualifiers  
 1. .217





```

QY 2245 TTTATTAAAAA 243 bp mRNA EST 27-AUG-1997
AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2276
|||||
Db 197 TTTATTAAAAA 5', mRNA sequence.
AAAAAAAAAAAAAAAAAAAAAAAAAAAA 228

RESULT 39
AA571665
LOCUS vm09e04.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
DEFINITION clone IMAGE:989694 5', mRNA sequence.
ACCESSION AA571665
VERSION AA571665
KEYWORDS AA571665.1 GI:2346594
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 243)
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
TITLE Unpublished (1998)
JOURNAL On May 9, 1995 this sequence version replaced gi:802596.
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:561974
Putative full length read
vector to vector length is 632
High quality sequence stop: 210.
FEATURES
Location/Qualifiers
1..243
/organism="Mus musculus"
/strain="BGD2 Fl/J"
/db_xref="taxon:10090"
/clone="IMAGE:989694"
/clone_lib="Knowles Solter mouse blastocyst B1"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/Note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGGTGACGCTGACCGCTTTTGT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
insert) and B3."
BASE COUNT 142 a 30 c 43 g 28 t
ORIGIN
Query Match 1.4%; Score 32; DB 35; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTAAAAA 243 bp mRNA EST 29-MAR-1999
AAAAAAAAAAAAAAAAAAAAAAAAAAAA 2276
|||||
Db 120 TTTATTAAAAA 243 bp mRNA EST
AAAAAAAAAAAAAAAAAAAAAAAAAAAA 151

RESULT 40
AU038878
LOCUS AU038878 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium
DEFINITION

Query Match 1.4%; Score 32; DB 44; Length 243;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2245 TTTATTAAAAA 247 bp mRNA EST 15-OCT-1999
XXXXXXXXXXXXXXXXXXXXXXXXXXXX 2276
|||||
Db 200 TTTATTAAAAA 247 bp mRNA EST
XXXXXXXXXXXXXXXXXXXXXXXXXXXX 231

RESULT 41
AU090738/c
LOCUS AU090738 247 bp mRNA EST 15-OCT-1999
DEFINITION xc95g05.x1 NCI-CGAP_Brn35 Homo sapiens cDNA clone IMAGE:2592056 3',
mRNA sequence.
ACCESSION AU090738
VERSION AU090738.1 GI:6048082
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 247)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
Unpublished (1998)
On Jul 9, 1999 this sequence version replaced gi:5435046.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```



Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bhrp/image/image.html](http://www-bio.llnl.gov/bhrp/image/image.html)

Seq primer: -40UP from Gibco

High quality sequence stop: 266.

Location/Qualifiers

## FEATURES

source

1. .269

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2441924"

/clone\_lib="NCI-CGAP\_Ot2"

/tissue\_type="moderately-differentiated endometrial

adenocarcinoma, 3 pooled tumors"

/lab\_host="DH10B"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.85 kb. Life Technologies catalog #:

11539-012"

BASE COUNT 70 a 36 c 53 g 110 t

## ORIGIN

## Query Match

Best Local Similarity 1.4%; Score 32; DB 62; Length 269;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA 2276

Db 43 TTTATTTAAAAA 12

## RESULT 45

AA749402/c

LOCUS

DEFINITION ny04h08.s1 NCI-CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1270815 3',

EST 18-FEB-1998

mrna sequence.

AA749402

VERSION AA749402.1 GI:2789360

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 272)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On Jan 14, 1998 this sequence version replaced gi:1877690.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bhrp/image/image.html](http://www-bio.llnl.gov/bhrp/image/image.html)

Insert Length: 1288 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from AmerSham

High quality sequence stop: 170.

Location/Qualifiers

source

1. .272

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1270815"

/clone\_lib="NCI-CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, 19b-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dt) primer

15'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTT-

3'). Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia) digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT713 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 25 c 30 g 119 t

## ORIGIN

Query Match 1.4%; Score 32; DB 38; Length 272;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2245 TTTATTTAAAAA 2276

Db 39 TTTATTTAAAAA 8

Search completed: April 6, 2000, 13:17:57

Job time: 21662 sec

-----

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 13:18:09 ; Search time 6421.81 Seconds  
(without alignments)  
79.372 Million cell updates/sec

Title: US-09-090-672B-9  
Perfect score: 135  
Sequence: 1 TTCTGACATGACTAAGAG.....AGGGACTGGACTAATAATGAC 135

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 4538634 seqs, 1887831982 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :		EST:*	
1:	em_est1:	*	
2:	em_est2:	*	
3:	em_est3:	*	
4:	em_est4:	*	
5:	em_est5:	*	
6:	em_est6:	*	
7:	em_est7:	*	
8:	em_est8:	*	
9:	em_est9:	*	
10:	em_est10:	*	
11:	em_est11:	*	
12:	em_est12:	*	
13:	em_est13:	*	
14:	em_est14:	*	
15:	em_est15:	*	
16:	em_est16:	*	
17:	em_est17:	*	
18:	em_est18:	*	
19:	em_est19:	*	
20:	gb_est1:	*	
21:	gb_est2:	*	
22:	gb_est3:	*	
23:	gb_est4:	*	
24:	gb_est5:	*	
25:	gb_est6:	*	
26:	gb_est7:	*	
27:	gb_est8:	*	
28:	gb_est9:	*	
29:	gb_est10:	*	
30:	gb_est11:	*	
31:	gb_est12:	*	
32:	gb_est13:	*	
33:	gb_est14:	*	
34:	gb_est15:	*	
35:	gb_est16:	*	
36:	gb_est17:	*	
37:	gb_est18:	*	
38:	gb_est19:	*	
39:	gb_est20:	*	
40:	gb_est21:	*	
41:	gb_est22:	*	
42:	gb_est23:	*	
43:	gb_est24:	*	

44:	gb_est25:	*	
45:	gb_est26:	*	
46:	gb_est27:	*	
47:	gb_est28:	*	
48:	gb_est29:	*	
49:	gb_est30:	*	
50:	gb_est31:	*	
51:	gb_est32:	*	
52:	em_est20:	*	
53:	em_est21:	*	
54:	em_est22:	*	
55:	em_est23:	*	
56:	em_est24:	*	
57:	em_est25:	*	
58:	em_est26:	*	
59:	gb_est33:	*	
60:	gb_est34:	*	
61:	gb_est35:	*	
62:	gb_est36:	*	
63:	gb_est37:	*	
64:	gb_est38:	*	
65:	em_est27:	*	
66:	em_est28:	*	
67:	em_est29:	*	
68:	em_est30:	*	
69:	gb_est39:	*	
70:	gb_est40:	*	
71:	gb_est41:	*	
72:	gb_est42:	*	
73:	gb_est43:	*	
74:	gb_est44:	*	
75:	em_est31:	*	
76:	em_est32:	*	
77:	em_est33:	*	
78:	em_est34:	*	
79:	gb_gss1:	*	
80:	gb_gss2:	*	
81:	gb_gss3:	*	
82:	gb_gss4:	*	
83:	em_gss1:	*	
84:	em_gss2:	*	
85:	em_gss3:	*	
86:	em_gss4:	*	
87:	gb_gss5:	*	
88:	gb_gss6:	*	
89:	gb_gss7:	*	
90:	gb_gss8:	*	
91:	gb_gss9:	*	
92:	em_gss5:	*	
93:	em_gss6:	*	
94:	em_gss7:	*	
95:	em_gss8:	*	
96:	em_gss9:	*	
97:	em_gss10:	*	
98:	em_gss11:	*	
99:	gb_gss10:	*	
100:	gb_gss11:	*	
101:	em_gss12:	*	
102:	gb_gss12:	*	
103:	gb_gss13:	*	
104:	gb_gss14:	*	
105:	gb_gss15:	*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		Query		Description	
No.	Score	Match	Length	ID	
C 1	18	13.3	278	41	AI007840
					AI007840 EST202291

```

c 2 18 13.3 364 42 A1145820 UI-R-BT0-
c 3 18 13.3 447 48 A1548491 UI-R-C3-t
c 4 18 13.3 498 43 A1100865 ES0210154
c 5 18 13.3 534 43 A1101330 EST210619
c 6 18 13.3 903 100 A0271832 nxbx0026D
c 7 17 12.6 175 81 B26945 T24A9TR TAM
c 8 17 12.6 340 90 A0072249 HS_3024_B
c 9 17 12.6 356 30 AA261914 zsl7d11.s
c 10 17 12.6 404 24 H73595 ysl1b08.s1
c 11 17 12.6 414 38 A0767354 nz81e12.s
c 12 17 12.6 427 28 AA126408 z19eh06.r
c 13 17 12.6 429 99 A0184537 HS_2203_A
c 14 17 12.6 435 30 AA262551 zsl7d11.r
c 15 17 12.6 447 30 AA228289 nc38c02.r
c 16 17 12.6 451 40 AA935611 op26g10.s
c 17 17 12.6 468 74 AW221275 EST297744
c 18 17 12.6 476 35 AA586533 nd46c09.s
c 19 17 12.6 483 99 A0240973 RPT11-64
c 20 17 12.6 484 102 A0402549 HS_5054_A
c 21 17 12.6 523 48 A1612731 tp11b09.x
c 22 17 12.6 528 69 AW106289 um25dl2.y
c 23 17 12.6 533 100 A0277193 CITBI-E1-
c 24 17 12.6 577 105 A0657152 Sheared D
c 25 17 12.6 601 81 B29489 F26010TFB I
c 26 17 12.6 618 100 A0278681 CITBI-E1-
c 27 17 12.6 620 91 A0157444 nbx0009B
c 28 17 12.6 764 90 A0078191 CIT-HSP-2
c 29 17 12.6 766 82 A0738602 HS_5381_B
c 30 17 12.6 931 81 B09381 T24A9-Sp6 T
c 31 16 11.9 96 27 AA052408 mb72h01.r
c 32 16 11.9 173 44 A1278373 qm63e02.x
c 33 16 11.9 195 49 AV011313 AV011313
c 34 16 11.9 213 35 AA570404 nk62g08.s
c 35 16 11.9 262 44 A1277304 qm67f07.x
c 36 16 11.9 263 105 A0663218 HS_5463_B
c 37 16 11.9 272 20 T02730 007793 gmbp
c 38 16 11.9 300 35 C34963 C34963 yuji
c 39 16 11.9 300 35 C36512 C36512 yuji
c 40 16 11.9 300 35 C51747 C51747 yuji
c 41 16 11.9 300 35 C52581 C52581 yuji
c 42 16 11.9 300 35 C52912 C52912 yuji
c 43 16 11.9 300 62 AV178626 AV178626
c 44 16 11.9 300 62 AV184617 AV184617
c 45 16 11.9 305 72 AV332351 AV332351

```

## ALIGNMENTS

```

RESULT 1
A1007840/c 278 bp mRNA EST 25-JAN-1999
LOCUS EST020291 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
DEFINITION RBRV56 3' end, mRNA sequence.
ACCESSION A1007840.1 GI:3221672
VERSION A1007840.1
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 278)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
JOURNAL Unpublished (1998)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900158.
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529

```

```

FEATURES             source
    source
    location/Qualifiers
    l..278
    /organism="Rattus sp."
    /db_xref="ATCC (inhost):2016112"
    /db_xref="taxon:10118"
    /clone="RBRV56"
    /clone_lib="Normalized rat brain, Bento Soares"
    /note="Organ: brain; Vector: p7T3Pac; Site_1: EcoRI;
    Site_2: NotI"
    BASE COUNT      86 a 70 c 53 g 59 t
    ORIGIN
        Query Match      13.3%; Score 18; DB 41; Length 278;
        Best Local Similarity 100.0%; Pred. No. 14;
        Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
        Qy 114 CAAGGGACTGGACTAAT 131
        Db 241 CAAGGGACTGGACTAAT 224
        |||||
        RESULT 2
        LOCUS A1145820/c 364 bp mRNA EST 05-JUL-1999
        DEFINITION UI-R-BT0-gi-12-0-UI.s1 UI-R-BT0 Rattus norvegicus cDNA clone
        A1145820
        ACCESSION A1145820
        VERSION A1145820.1 GI:3667619
        KEYWORDS EST.
        SOURCE Norway rat.
        ORGANISM Rattus norvegicus
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
        Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
        REFERENCE 1 (bases 1 to 364)
        AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
        TITLE Normalization and subtraction: two approaches to facilitate gene
        discovery
        JOURNAL Genome Res. 6 (9), 791-806 (1996)
        MEDLINE 97044477
        COMMENT On Jan 17, 1998 this sequence version replaced gi:2152861.
        Contact: Soares, MB
        Program for Rat Gene Discovery and Mapping
        University of Iowa
        451 Eckstein Medical Research Building Iowa City, IA 52242, USA
        Tel: 319 335 8250
        Fax: 319 335 9565
        Email: msoares@blue.weeg.uiowa.edu
        The sequence tag present in the cDNA between the NotI site and the
        oligo-dT track served to identify it as a clone from the normalized
        adult NR6thalamus library. cDNA Library Preparation: M. Fatima
        Bonaldo, Ph.D. Clone distribution: clones will be available through the
        Research Genetics This clone is also available through the
        I.M.A.G.E. Consortium at LLNL (infoimage.llnl.gov). IMAGE
        ID-1788870 The following repetitive elements were found in this
        cDNA sequence: 1-27, >AT-rich#Low_complexity
        Seq primer: M13 Forward
        POLYA-No. Location/Qualifiers
        l..364
        /organism="Rattus norvegicus"
        /strain="Sprague-Dawley"
        /db_xref="taxon:10116"
        /clone="UI-R-BT0-gi-g-12-0-UI"
        /clone_lib="UI-R-BT0"
        /dev_stage="adult"
        /lab_host="DH10B (Life Technologies)"
        /note="Vector: p7T3B-Pac (Pharmacia) with a modified
        polylinker; Site_1: Not I; Site_2: Eco RI; This library
        (UI-R-BT0) consists of a mixture of individually tagged

```

normalized libraries constructed from rat hippocampus, thalamus, mid-brain, medulla, corpus striatum, cerebral cortex and testis. The tag used to identify the source tissue is a string of 3-6 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. This library was then subcloned using a driver consisting of a mixture of all clones from UI-R-A0, UI-R-A1, UI-R-E0, UI-R-E1, UI-R-C0, UI-R-C1, UI-R-C2 and UI-R-C2p."

BASE COUNT 106 a 89 c 66 g 103 t  
ORIGIN

Query Match 13.3%; Score 18; DB 42; Length 364;

Best Local Similarity 100.0%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGACTGGACTAAAT 131  
|||||  
Db 258 CAAGGACTGGACTAAAT 241

## RESULT 3

AI548491/c

LOCUS

DEFINITION

AI548491

AI548491

AI548491.1

GI:4465979

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AI548491 447 bp mRNA EST 22-MAR-1999  
UI-R-C3-to-a-04-0-UI.s1 UI-R-C3 Rattus norvegicus cDNA clone  
UI-R-C3-to-a-04-0-UI 3', mRNA sequence.  
AI548491  
AI548491.1 GI:4465979  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
On Jun 5, 1998 this sequence version replaced gi:3188189.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.wesg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized brain library cDNA Library Preparation: M.B. Soares Lab  
Clone distribution: clones will be available through Research  
Genetics (www.resgen.com) The following repetitive elements were  
found in this cDNA sequence: 1-34, >AT-rich#low\_complexity  
Seq primer: M13 Forward.

## FEATURES

source

Location/Qualifiers

1..447  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-C3-to-a-04-0-UI"  
/clone\_lib="UI-R-C3"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"

/note="Vector: pMT3D-Pac (Pharmacia) with a modified  
polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-C3  
library is a subtracted library of a series, ultimately  
derived from a mixture of individually tagged normalized  
libraries from rat placenta, adult lung, brain, liver,

kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day  
embryos, after a series of subtractions to reduce the  
representation of cDNAs from which ESTs had already been  
generated. The following serially subtracted libraries  
were generated in this process: UI-R-C3, UI-R-C2p,  
UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of  
3-5 nucleotides present between the Not I site and the  
oligo-dT track which allows identification of the library  
of origin of a clone within the mixture. The subtracted  
library (UI-R-C3) was constructed as follows: PCR amplified  
cDNA inserts from UI-R-C2p clones from which 3' ESTs had  
been derived was used as a driver in a hybridization with  
the UI-R-C2p library in the form of single-stranded  
circles. The remaining single-stranded circles (subtracted  
library) was purified by hydroxyapatite column  
chromatography, converted to double-stranded circles and  
electroporated into DH10B bacteria (Life Technologies) to  
generate the UI-R-C3 library. This procedure has been  
previously described (Bonaldo, Lennon and Soares, Genome  
Research 6:791-806, 1996)"

BASE COUNT 134 a 110 c 81 g 122 t  
ORIGIN

Query Match 13.3%; Score 18; DB 48; Length 447;

Best Local Similarity 100.0%; Pred. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGACTGGACTAAAT 131  
|||||  
Db 265 CAAGGACTGGACTAAAT 248

## RESULT 4

AI100865/c

LOCUS

DEFINITION

AI100865

AI100865

AI100865.1

GI:3705892

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AI100865 498 bp mRNA EST 31-JAN-1999  
EST210154 Normalized rat brain, Bento Soares Rattus sp. cDNA clone  
RBBB71 3' end, mRNA sequence.  
AI100865  
AI100865.1 GI:3705892  
EST.  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,  
Kerlavage, A.R. and Adams, M.D.  
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat  
Gene Index  
Unpublished (1998)  
Other ESTs: TC54489  
Contact: Lee, NH  
ATCC  
The Institute for Genomic Research  
9712, Medical Center Drive, Rockville, MD 20850, USA  
Tel: (301)-838-3529  
Fax: (301)-838-0208  
Email: nhlee@igr.org  
Seq primer: M13-21.  
Location/Qualifiers

FEATURES  
source

1..498  
/organism="Rattus sp."  
/db\_xref="ATCC (inhost):2023824"  
/db\_xref="taxon:10118"  
/clone="RBBB71"  
/clone\_lib="Normalized rat brain, Bento Soares"  
/note="Organ: brain; vector: pMT3D-Pac; Site.1: EcoRI;  
Site.2: NotI"

BASE COUNT 149 a 120 c 108 g 121 t  
ORIGIN

```

Query Match      13.3%; Score 18; DB 43; Length 498;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGGACTGGACTAAAT 131
      |||||
DB 247 CAAGGGACTGGACTAAAT 230

RESULT 5
AL101330/c
LOCUS      534 bp mRNA EST 31-JAN-1999
DEFINITION EST210619 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
            RBRBL55 3' end, mRNA sequence.
ACCESSION  AL101330
VERSION     AL101330
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 534)
AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kertlavage,A.R. and Adams,M.D.
TITLE     Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
            Gene Index
JOURNAL    Unpublished (1998)
COMMENT    Other_ESTs: TC54489
            Contact: Lee, NH
            ATCC

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES             source
    source
    1..534
        /organism="Rattus sp."
        /db_xref="ATCC (inhost):2029307"
        /db_xref="taxon:10118"
        /clone="RBRBL55"
        /clone_lib="Normalized rat brain, Bento Soares"
        /note="Organ: brain; Vector: pT7T3Pac; Site_1: EcoRI;
        Site_2: NotI"
BASE COUNT  155 a 130 c 117 g 132 t

Query Match      13.3%; Score 18; DB 43; Length 534;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 CAAGGGACTGGACTAAAT 131
      |||||
DB 244 CAAGGGACTGGACTAAAT 227

RESULT 6
AQ271832/c
LOCUS      903 bp DNA GSS 03-NOV-1998
DEFINITION nbxb0026D08f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0026D08f, genomic survey sequence.
ACCESSION  AQ271832
VERSION     AQ271832.1 GI:3825147
KEYWORDS   GSS.
SOURCE     Oryza sativa.
            Oryza sativa.
ORGANISM   Oryza sativa.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE  1 (bases 1 to 903)
AUTHORS   Wing,R.A. and Dean,R.A.

```

```

TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL    Unpublished (1998)
COMMENT    Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 88.
            Location/Qualifiers
                1..903
                    /organism="Oryza sativa"
                    /strain="Japonica"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:4530"
                    /clone="nbxb0026D08f"
                    /clone_lib="CUGI Rice BAC Library"
                    /tissue_type="Leaf"
                    /lab_host="E. coli DH10B"
                    /note="Vector: pBelosBAC11; Site_1: HindIII; Site_2:
                    HindIII; Rice is one of two most popular grains in the
                    world. Half of the world population especially those
                    inhabiting highly populated areas of the humid tropics
                    and subtropics, rely on rice as their primary source of
                    carbohydrate. Monocytledonous rice is a diploid plant
                    (2n=24) with a haploid genome equivalent of 431 Mbp
                    (Arumuganathan and Earle, 1991). The relatively small
                    genome of rice, three times larger than that of
                    Arabidopsis, makes it suitable for genomic studies. In
                    order to facilitate positional cloning, physical mapping
                    and genome sequencing of rice, we have constructed a BAC
                    library from Oryza sativa, Nipponbare variety. The
                    library contains 36,864 clones with an average insert size
                    of 128.5 Kb providing 10.9 haploid genome equivalents.
                    The deep coverage allows the isolation a particular
                    sequence with a probability of 99.9 %. Two high density
                    filters, each containing 18,432 clones (doubly spotted),
                    represent the whole library for colony screening."
BASE COUNT  192 a 183 c 89 g 419 t 20 others
ORIGIN

Query Match      13.3%; Score 18; DB 100; Length 903;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 AATGAGTAAAGAGAAAGA 25
      |||||
DB 750 AATGAGTAAAGAGAAAGA 733

RESULT 7
B26945
LOCUS      175 bp DNA GSS 13-OCT-1997
DEFINITION T24A9TR TAMU Arabidopsis thaliana genomic clone T24A9, genomic
            survey sequence.
ACCESSION  B26945
VERSION     B26945.1 GI:2512911
KEYWORDS   GSS.
SOURCE     thale cress.
            Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE  1 (bases 1 to 175)
AUTHORS   Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and
            Venter,J.C.
            Use of a BAC End Sequence Database To Identify Minimal Overlaps for
            Arabidopsis Genomic Sequencing
TITLE

```



JOURNAL  
COMMENT

Unpublished (1997)  
Other.GSSs: T2A9TfB  
Contact: Steve Rounsley  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: rounsley@tigr.org  
Seg primer: M13 Reverse  
Class: BAC ends  
High quality sequence stop: 175.

FEATURES  
source

Location/Qualifiers  
1..175  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T2A9"  
/clone\_lib="TAMU"  
/sex="hermaphrodite"  
/note="Vector: BelOBACII; Site\_1: HindIII; Site\_2:  
HindIII; Produced by Rod Wing"  
60 a 31 c 25 g 59 t

BASE COUNT  
ORIGIN

Query Match 12.6%; Score 17; DB 81; Length 175;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ATTAAGATTATCATAG 60  
|||||

Db 116 ATTAAGATTATCATAG 132  
|||||

RESULT 8  
AQ072249  
LOCUS

DEFINITION  
sapiens genomic clone Plater-3024 Col-1 Row-N, genomic survey  
sequence.

ACCESSION  
AQ072249  
VERSION  
AQ072249.1 GI:3391098  
KEYWORDS  
GSS.  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 340)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and  
Hood,L.

TITLE  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL  
MEDLINE  
COMMENT  
99380589  
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887

Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3024 row: N column: 1

Class: BAC ends  
High quality sequence stop: 340.

## FEATURES

## source

Location/Qualifiers  
1..340  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plater-3024 Col-1 Row-N"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"  
/note="Organ: sperm; Vector: pBelOBACII; BAC Clones in  
E-Coli DH10B"

BASE COUNT 127 a 44 c 58 g 111 t  
ORIGIN

Query Match 12.6%; Score 17; DB 90; Length 340;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ATTATCATAGAGCAAT 66  
|||||

Db 64 ATTATCATAGAGCAAT 80  
|||||

## RESULT 9

## AA261914/c

## LOCUS

## DEFINITION

AA261914 356 bp mRNA EST 13-AUG-1997

ZS17/dll.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:685461 3',  
mRNA sequence.

ACCESSION  
AA261914

VERSION  
AA261914.1 GI:1897748

KEYWORDS  
EST.

SOURCE  
human.

ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 356)

AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
Unpublished (1997)

COMMENT  
On May 8, 1995 this sequence version replaced gi:800902.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 436-1550

Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: 41ml3 fwd, ET from Amersham.

FEATURES  
source

Location/Qualifiers  
1..356  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:685461"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
15'-TGTTACCAATCTCAAGTGGAGCGCGCTCATTTTTTTTTTTT-  
3'. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 117 c 88 g 66 t  
ORIGIN

Query Match 12.6%; Score 17; DB 30; Length 356;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCTTGCCCTTTGGTTAT 45  
|||||

Db 354 TCTTGCCCTTTGGTTAT 338  
|||||



KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 427)  
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
Roehlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397564.  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28m13 rev2 from Amersham  
High quality sequence stop: 277.  
Location/Qualifiers  
1. 427  
/organism="Homo sapiens"  
/db\_xref="GDB:3844042"  
/db\_xref="taxon:9606"  
/clone="IMAGE:511547"  
/lab\_host="Stratagene colon (#937204)"  
/note="Organ: colon; Vector: pBluescript SK-; site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. T-84 colonic epithelial cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGCGCAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 114 a 121 c 114 g 76 t 2 others  
ORIGIN

Query Match 12.6%; Score 17; DB 28; Length 427;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 AGGCTCTGCCCTTGG 41  
Db 354 AGGCTCTGCCCTTGG 338  
|||||

RESULT 13  
AQ184537/c  
LOCUS  
DEFINITION HS\_2205\_AL\_G01\_MF CIT Approved Human Genomic Sperm Library D Homo  
sapiens genomic clone Plate-2205 Col-1 Row-M, genomic survey  
sequence.  
ACCESSION AQ184537  
VERSION AQ184537  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 429)  
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,  
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and  
Hood, L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 2205 row: M column: 1  
Class: BAC ends  
High quality sequence stop: 429.  
Location/Qualifiers  
1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate-2205 Col-1 Row-M"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelosBAC11; BAC Clones in  
E-Coli DH10B"  
BASE COUNT 126 a 82 c 79 g 139 t 3 others  
ORIGIN

Query Match 12.6%; Score 17; DB 99; Length 429;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 AGAAGAAAGAGGCTT 32  
Db 155 AGAAGAAAGAGGCTT 139  
|||||

RESULT 14  
AA262551  
LOCUS  
DEFINITION zsl7d11.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:685461 5',  
mrna sequence.  
ACCESSION AA262551  
VERSION AA262551.1 GI:1897913  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 435)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 8, 1995 this sequence version replaced gi:801104.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Putative full length read  
The vector to vector length is 436  
Seq primer: -28m13 rev2 ET from Amersham.  
Location/Qualifiers  
1. 435  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:685461"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),

Provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTT-3']. Double-stranded CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 83 a 112 c 136 g 104 t  
 ORIGIN

Query Match 12.6%; Score 17; DB 30; Length 435;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TCTTGCCCTTGGTTAT 45  
 Db 72 TCTTGCCCTTGGTTAT 88

RESULT 15  
 AA228289  
 LOCUS 447 bp mRNA EST 21-AUG-1997  
 DEFINITION nc38c02.r1 NCI-CGAP\_P2 Homo sapiens CDNA clone IMAGE:1010402  
 similar to contains Alu repetitive element; contains element PTR5  
 repetitive element ; mRNA sequence.

ACCESSION AA228289.1 GI:1849850  
 VERSION AA228289  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 447)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393543.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui,  
 M.D., Michael Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 DNA sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 505 Std Error: 0.00  
 Seq primer: -28ml3 rev1 ET from Amersham  
 High quality sequence stop: 409.  
 Location/Qualifiers  
 1. 447  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1010402"  
 /clone\_lib="NCI-CGAP\_P2"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
 strand cDNA was primed with oligo(dT)17 on 50 ng Of  
 DNase-treated, total cellular RNA obtained from  
 5,000-10,000 microdissected preneoplastic cells  
 histologically determined to be prostatic intraepithelial  
 neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
 ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
 cDNA with an adaptor-specific primer, and the resulting

PCR product subcloned into pAMP10 by the UDG-cloning  
 method (Life Technologies). Average insert size is 600  
 bp. NOTE: Not directionally cloned. This library was  
 constructed by David Krizman."

BASE COUNT 128 a 96 c 111 g 112 t  
 ORIGIN

Query Match 12.6%; Score 17; DB 30; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GACAAATGAGTAAGAAGA 21  
 Db 61 GACAAATGAGTAAGAAGA 77

RESULT 16  
 AA935611  
 LOCUS 451 bp mRNA EST 07-JUL-1998  
 DEFINITION OP26g10.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens CDNA clone  
 IMAGE:1578018 3' similar to WP:C14C11.3 CE08822 ; mRNA sequence.

ACCESSION AA935611  
 VERSION AA935611.1 GI:3092768  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 451)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900158.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 1658 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 148.

FEATURES  
 source  
 1. 451  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1578018"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NBHL19W, testis NHT, and B-cell  
 NCI-CGAP-GCBI) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Fatima Bonaldo."  
 106 a 121 c 126 g 98 t

BASE COUNT 106 a 121 c 126 g 98 t  
 ORIGIN

Query Match 12.6%; Score 17; DB 40; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAATGAGTAAGAAGA 23  
 Db 241 CAATGAGTAAGAAGA 257

```

RESULT 17
AW221275
LOCUS
DEFINITION EST297744 tomato fruit mature green, TAMU Lycopersicon esculentum
cdna clone cLEF13N5, mRNA sequence.
ACCESSION AW221275
VERSION AW221275.1 GI:6532959
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Potatoe; Lycopersicon.
1 (bases 1 to 468)
Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F.,
Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S.,
Ronning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and
Giovannoni, J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
On Mar 10, 1998 this sequence version replaced gi:2948814.
Contact: David Frisch
Clemson University
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu
5 prime sequence.
FEATURES
source
1..468
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEF13N5"
/clone_lib="tomato fruit mature green, TAMU"
/tissue_type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="vector: phuescript SK(-); site_1: EcoRI; site_2:
XhoI; cLEF - Fruit were tagged at the 1cm stage and
harvested 3-5 days prior to ripening. Fruit were cut in
half to verify the seeds were indeed 'immature' and the
seeds and locules were discarded prior to freezing the
pericarp"
BASE COUNT 136 a 108 c 92 g 132 t
ORIGIN
Query Match 12.6%; Score 17; DB 74; Length 468;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 ACCAGCACAGAGATTAG 103
Db 178 ACCAGCACAGAGATTAG 194
|||||
RESULT 18
AA586533 476 bp mRNA EST 12-SEP-1997
LOCUS nd46c09.s1 NCI_CGAP_Col Homo sapiens cDNA clone IMAGE:816016.3'
DEFINITION similar to gb:X01677 Glyceroldehyde 3-phosphate dehydrogenase,
LIVER (HUMAN);, mRNA sequence.
ACCESSION AA586533
VERSION AA586533.1 GI:2397347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
COMMENT Tumor Gene Index
Unpublished (1997)
On May 9, 1995 this sequence version replaced gi:802594.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Seq primer: -21m13 forward (Amersham)
High quality sequence stop: 1.
FEATURES
source
1..476
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:6037592"
/db_xref="taxon:9606"
/clone="IMAGE:816016"
/clone_lib="NCI_CGAP_Col"
/tissue_type="bulk tumor"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-Sport2; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dn."
BASE COUNT 106 a 114 c 134 g 122 t
ORIGIN
Query Match 12.6%; Score 17; DB 35; Length 476;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 AAATCGTGTATACCA 90
Db 253 AAATCGTGTATACCA 269
|||||
RESULT 19
AA240973 483 bp DNA GSS 21-APR-1999
LOCUS RPC11-64B24.TK RPC1-11 Homo sapiens genomic clone RPC1-11-64B24,
DEFINITION genomic survey sequence.
ACCESSION AA240973
VERSION AA240973.1 GI:3668557
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 483)
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)
Other GSSs: RPC11-64B24.TJ
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are derived from the human BAC library RPC1-11. For BAC
library availability, please contact Pieter de Jong

```

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: 17  
Class: BAC ends.

#### FEATURES

source  
Location/Qualifiers  
1..483  
/organism="Homo sapiens"  
/db\_xref="GDB:7524239"  
/db\_xref="taxon:9606"  
/clone\_lib="RPCI-11-64824"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"  
153 a 96 c 92 g 142 t

#### BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; DB 99; Length 483;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 TTATCATAGACCAATAA 58

Db 172 TTATCATAGACCAATAA 156

#### RESULT 20

AQ402549/c

LOCUS

DEFINITION

HS\_5054\_A1\_E06\_T7A RPCI-11 Human Male BAC Library Homo sapiens

genomic clone Plate-630 Col-11 Row-I, genomic survey sequence.

ACCESSION

AQ402549

VERSION

AQ402549.1 GI:4413461

KEYWORDS

GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 484)

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))

/sex="male"  
/note="Vector: pBACE3.6; Genomic sequence of BAC ends"  
144 a 101 c 105 g 128 t 6 others

#### BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; OB 102; Length 484;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 ACAATAAACCAAGGA 120

Db 74 ACAATAAACCAAGGA 58

#### RESULT 21

AI612731/c

LOCUS

DEFINITION

tp11b09.x1 NCI-CGAP\_Gas4 Homo sapiens cDNA clone IMAGE:218749 3'

similar to gb:X06617 40S RIBOSOMAL PROTEIN S11 (HUMAN);, mRNA

sequence.

ACCESSION

AI612731

VERSION

AI612731.1 GI:4621898

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 523)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

On May 18, 1998 this sequence version replaced gi:3136494.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.O., Ph.O., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40UP from Gibco

High quality sequence stop: 67.

FEATURES

source

Location/Qualifiers

1..523

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:218749"

/clone\_lib="NCI-CGAP\_Gas4"

/tissue\_type="poorly differentiated adenocarcinoma with

signet ring cell features"

/lab\_host="DH108"

/note="Organ: stomach; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo #1.  
Average insert size 1.69 kb. Life Technologies catalog #:  
11549-011"

#### BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; OB 48; Length 523;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 AAGAAAGAGGCTTGC 34

Db 507 AAGAAAGAGGCTTGC 491

#### FEATURES

source

Location/Qualifiers

1..484

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-630 Col-11 Row-I"

/clone\_lib="RPCI-11 Human Male BAC Library"

```

RESULT 22
AW106269
LOCUS
DEFINITION
IMAGE:2225591 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
On Dec 20, 1995 this sequence version replaced gi:1134183.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLMN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1005787
Seq primer: custom primer used
High quality sequence stop: 443.
FEATURES
Location/Qualifiers
1..528
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2225591"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGCTGTC);
Site_2: DraIII (CACCATGG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGCGCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGGCGCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site
CACCATGG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTTAAGAGCTGCG and 3' end primer
CGACTCGAGCTCGAGCACA."
BASE COUNT 155 a 107 c 112 g 154 t
ORIGIN
Query Match 12.6%; Score 17; DB 69; Length 528;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 GTAAGAAGAGAGGCT 29
Db 369 GTAAGAAGAGAGGCT 385
RESULT 23
AQ277193
LOCUS
DEFINITION
IMAGE:2522C17.TF C17B1-E1 Homo sapiens genomic clone 2522C17,
genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.
Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building (1998)
Unpublished (1998)
Other_GSSs: CITB1-E1-2522C17.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES
Location/Qualifiers
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2522C17"
/clone_lib="CITB1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
CalTech Human BAC Library D"
BASE COUNT 143 a 136 c 176 t
ORIGIN
Query Match 12.6%; Score 17; DB 100; Length 533;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 CAATGAGTAGAAGAAA 23
Db 397 CAATGAGTAGAAGAAA 413
RESULT 24
AQ657152/c
LOCUS
DEFINITION
IMAGE:12F21.TF Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-12F21, genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE
AUTHORS
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,
Donelson, J., Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

```

Tel: 301 838 0200  
Fax: 301 838 0208

Email: nelsayed@tigr.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mdb/tbdb/>.  
seq primer: M13-Forward  
Class: Shotgun.

#### FEATURES

Location/Qualifiers

1..577  
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone="Sheared DNA-12F21"  
/clone\_lib="Sheared DNA"  
/note="Vector: pUC18; Site\_1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 192 a 97 c 168 g 120 t  
ORIGIN

Query Match 12.6%; Score 17; DB 105; Length 577;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 48 AGATTATCATAGAGCA 64

Db 87 AGATTATCATAGAGCA 71

#### RESULT 25

B29489

LOCUS

DEFINITION F26010TFB IGF Arabidopsis thaliana genomic clone F26010, genomic survey sequence.

ACCESSION B29489

VERSION B29489.1 GI:2515455

KEYWORDS GSS.

SOURCE

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 601)

Rounsley,S.D., Kelley,J.M., Field,C.E., Craven,M.B., Adams,M.D. and Venter,J.C.

Use of a BAC End Sequence Database To Identify Minimal Overlaps for Arabidopsis Genomic Sequencing

JOURNAL

COMMENT

Other GSSs: F26010TRC

Contact: Steve Rounsley

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: rounsley@tigr.org

Seq primer: M13-21

Class: BAC ends

High quality sequence stop: 601.

#### FEATURES

Location/Qualifiers

1..601

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db\_xref="taxon:3702"

/clone="F26010"

/clone\_lib="IGF"

/sex="hermaphrodite"

/note="Vector: pBelBACII; Site\_1: EcoRI; Site\_2: EcoRI; Produced by Thomas Altmann"

213 a 94 c 105 g 189 t

#### BASE COUNT

ORIGIN

Query Match 12.6%; Score 17; DB 81; Length 601;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 44 ATTAAGATTATCATAG 60

Db 581 ATTAAGATTATCATAG 597

#### RESULT 26

LOCUS

DEFINITION AQ278681 618 bp DNA GSS 22-NOV-1998

genomic survey sequence.

ACCESSION AQ278681

VERSION AQ278681.1 GI:3904649

KEYWORDS GSS.

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 618)

Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,

Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and

Venter,J.C.

Use of a random human BAC End Sequence Database for Sequence-Ready

Map Building

Unpublished (1998)

Other\_GSSs: CITBI-El-251685.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com).

end search page:

[http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html).

Seq primer: M13 Reverse

Class: BAC ends.

#### FEATURES

Location/Qualifiers

1..618

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="251685"

/clone\_lib="CITBI-El"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBelBACII; Site\_1: EcoRI; Site\_2: EcoRI;

Caltech Human BAC Library D"

173 a 153 c 127 g 165 t

#### BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 AGTAAGAAAGAGGG 28

|||||



```

Db 330 AGTAAGAGAGAGAGGG 346

RESULT 27
AQ157444          620 bp      DNA      GSS      12-SEP-1998
LOCUS
DEFINITION
  nbxb0009B13r CUGI Rice BAC Library Oryza sativa genomic clone
  nbxb0009B13r, genomic survey sequence.
ACCESSION
  AQ157444
VERSION
  nbxb0009B13r
KEYWORDS
  GSS.
SOURCE
  Oryza sativa.
  Oryza sativa.
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
  Poaceae; Oryza.
  1 (bases 1 to 620)
REFERENCE
  Wing, R.A. and Dean, R.A.
  A BAC End Sequencing Framework to Sequence the Rice Genome
  Unpublished (1998)
AUTHORS
  Contact: Wing RA
  Clemson University Genomics Institute
  100 Jordan Hall, Clemson, SC 29634, USA
  Tel: 864 656 7288
  Fax: 864 656 4293
  Email: rwing@clemson.edu
  Seq primer: GGAACAGCTATGACCATG
  Class: BAC ends
  High quality sequence start: 54
  High quality sequence stop: 294.
FEATURES
  Location/Qualifiers
    1..620
    /organism="Oryza sativa"
    /strain="Japonica"
    /cultivar="Nipponbare"
    /db_xref="taxon:4530"
    /clone="nbxb0009B13r"
    /clone_lib="CUGI Rice BAC Library"
    /tissue_type="Leaf"
    /lab_host="E. coli DH10B"
    /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
    HindIII; Rice is one of two most popular grains in the
    world. Half of the world population especially those
    inhabiting highly populated areas of the humid tropics
    and subtropics, rely on rice as their primary source of
    carbohydrate. Monocotyledonous rice is a diploid plant
    (2n=24) with a haploid genome equivalent of 431 Mbp
    (Arumuganathan and Earle, 1991). The relatively small
    genome of rice, three times larger than that of
    Arabidopsis, makes it suitable for genomic studies. In
    order to facilitate positional cloning, physical mapping
    and genome sequencing of rice, we have constructed a BAC
    library from Oryza sativa, Nipponbare variety. The
    library contains 36,864 clones with an average insert size
    of 128.5 kb providing 10.9 haploid genome equivalents.
    The deep coverage allows the isolation a particular
    sequence with a probability of 99.9 %. Two high density
    filters, each containing 18,432 clones (doubly spotted),
    represent the whole library for colony screening."
  BASE COUNT      174 a 99 c 104 g 243 t
  ORIGIN

  Query Match          12.6%; Score 17; DB 91; Length 620;
  Best Local Similarity 100.0%; Pred. No. 38;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 29 TCTTGCCCTTGTTAT 45
  |||||||
  Db 560 TCTTGCCCTTGTTAT 576

RESULT 28
AQ078191          764 bp      DNA      GSS      20-AUG-1998
LOCUS
DEFINITION
  CIT-HSP-2360M3.TR CIT-HSP Homo sapiens genomic clone 2360M3,
  genomic survey sequence.
ACCESSION
  AQ078191
VERSION
  AQ078191.1 GI:3439375
KEYWORDS
  GSS.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 764)
REFERENCE
  Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
  Venter,J.C.
  Use of a random human BAC End Sequence Database for Sequence-Ready
  Map Building
  Unpublished (1998)
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
  Seq primer: M13 Reverse
  Class: BAC ends.
FEATURES
  Location/Qualifiers
    1..764
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="2360M3"
    /clone_lib="CIT-HSP"
    /sex="Male"
    /cell_type="Sperm"
    /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
    HindIII"
  BASE COUNT      241 a 174 c 143 g 206 t
  ORIGIN

  Query Match          12.6%; Score 17; DB 90; Length 764;
  Best Local Similarity 100.0%; Pred. No. 36;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

  Qy 52 TTATCATAGAGCAATAA 68
  |||||||
  Db 82 TTATCATAGAGCAATAA 66

RESULT 29
AQ0738602         766 bp      DNA      GSS      16-JUL-1999
LOCUS
DEFINITION
  HS_5381.B1.B07.SP6.RPCI-11 Human Male BAC Library Homo sapiens
  genomic clone Plate=957 Col=13 Row=D, genomic survey sequence.
ACCESSION
  AQ0738602
VERSION
  AQ0738602.1 GI:5516124
KEYWORDS
  GSS.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 766)
REFERENCE
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
  Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  JOURNAL

```

```

MEDLINE
COMMENT
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallaceu.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact pletier de Jong
(pletier@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (Info@resgen.com). BAC end Web Server:
http://www.hsc.washington.edu
Plate: 957 row: D column: 13
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 766.
FEATURES
Source
1..766
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 160 a 251 c 179 g 176 t
ORIGIN
11 GAGTAAAGAGAGAGG 27
|||||
724 GAGTAAAGAGAGAGG 708
|||||

Query Match 12.6%; Score 17; DB 82; Length 766;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 GAGTAAAGAGAGAGG 27
|||||
724 GAGTAAAGAGAGAGG 708
|||||

RESULT 30
B09381
LOCUS
DEFINITION
T24A9-SP6 TAMU Arabidopsis thaliana genomic clone T24A9, genomic
survey sequence.
ACCESSION B09381
VERSION B09381.1 GI:2090501
KEYWORDS
SOURCE
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE
1 (bases 1 to 931)
Feng J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and
Ecker, J.
BAC End Sequences at ATGC
Unpublished (1997)
Other_GSSs: T24A9-T7
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 85
High quality sequence stop: 307.
FEATURES
Source
1..931
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="T24A9"
/sex="hermaphrodite"
/note="Vector: BelOBAC11; Site_1: HindIII; Site_2:
HindIII; Produced by Rod Wing"
BASE COUNT 283 a 190 c 154 g 298 t 6 others
ORIGIN
44 ATTAAGATTATCATAG 60
|||||
189 ATTAAGATTATCATAG 205
|||||

Query Match 12.6%; Score 17; DB 81; Length 931;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 ATTAAGATTATCATAG 60
|||||
189 ATTAAGATTATCATAG 205
|||||

RESULT 31
AA052408
LOCUS
DEFINITION
mb72h01.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
IMAGE:334993 5' similar to gb:M21389 KERATIN, TYPE II CYTOSKELETAL
5 (HUMAN); mRNA sequence.
ACCESSION AA052408
VERSION AA052408.1 GI:1543263
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 96)
Marra M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMNI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:216393
Trace considered overall poor quality
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 1.
FEATURES
Source
1..96
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="334993"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: p7T3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAGTGGAGCGCGCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Donaldso. RNA was kindly provided by
```

```

BASE COUNT      22 a      26 c      26 g      22 t
ORIGIN
    Query Match      11.9%; Score 16; DB 27; Length 96;
    Best Local Similarity 100.0%; Pred. No. 2.3e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CAATGAGTAAGAGAA 22      29-JAN-1999
    |||||||
Db 70 CAATGAGTAAGAGAA 85

RESULT 32
AI278373
LOCUS      qm63e02.x1 Soares_placenta_8to9weeks_2NbHP8to9w Homo sapiens cDNA
DEFINITION clone IMAGE:1893434 3', mRNA sequence.
ACCESSION  AI278373
VERSION     AI278373
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 173)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (infoimage.llnl.gov) for further information.
            Insert Length: 1154 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 150.
FEATURES             Location/Qualifiers
     source           1..173
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1893434"
                     /clone_lib="Soares_placenta_8to9weeks_2NbHP8to9w"
                     /dev_stage="two placentae: one from 8 weeks and another
                     from 9 weeks post conception"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: p7T73p (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGCGCGCGATTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified p7T73 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M. Fatima Bonaldo."
BASE COUNT      52 a      32 c      22 g      67 t
ORIGIN
    Query Match      11.9%; Score 16; DB 44; Length 173;
    Best Local Similarity 100.0%; Pred. No. 1.9e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GATTATCATAGACGA 64
    |||||||
Db 88 GATTATCATAGACGA 103

RESULT 33
AV011313/c
LOCUS      AV011313.1 GI:4788300
DEFINITION AV011313 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
            clone 1110031J17, mRNA sequence.
ACCESSION  AV011313
VERSION     AV011313.1
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 195)
            Carlini, P., Akiyama, J., Fukuda, S., Konno, H., Itoh, M., Aizawa, K.,
            Akahira, S., Shibata, K., Ozawa, Y., Hara, A., Hayatsu, N., Hori, F.,
            Ishikawa, T., Ito, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima,
            Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C., Sato, K., Shibata,
            Y., Shigemoto, F., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
            Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
            Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
            RIKEN Mouse ESTs
            Unpublished (1999)
            On Jun 5, 1998 this sequence version replaced gi:3188818.
COMMENT    Contact: Chie Owa
            Genome Science Laboratory
            RIKEN
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-9145
            Fax: 81-298-36-9098
            Email: genome-res@rtc.riken.go.jp
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length cDNA
            (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
            Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source           1..195
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1110031J17"
                     /clone_lib="Mus musculus 18-day embryo C57BL/6J"
                     /sex="mixed"
                     /dev_stage="18-day embryo"
BASE COUNT      50 a      50 c      33 g      62 t
ORIGIN
    Query Match      11.9%; Score 16; DB 49; Length 195;
    Best Local Similarity 100.0%; Pred. No. 1.9e+02;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 TGAGTAAGAAGAAAGA 25
    |||||||
Db 44 TGAGTAAGAAGAAAGA 29

RESULT 34
AA570404
LOCUS      AA570404.1 NCICGAP_Sch1 Homo sapiens cDNA clone IMAGE:1018142 3',
DEFINITION mRNA sequence.
ACCESSION  AA570404
VERSION     AA570404.1
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 213)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index

```

JOURNAL  
COMMENT

Unpublished (1997)  
On Nov 29, 1993 this sequence version replaced gi:637706.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc., David B. Krizman,  
Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1191 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 144.

FEATURES  
Source

1. .213  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_Schl"  
/tissue\_type="Schwannoma tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; site\_1: EcoRI; site\_2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT. Two  
pooled bulk Schwannoma tumors. 5' adaptor sequence: 5'  
CGATCGCACAG 3' 3' adaptor sequence: 5'  
CGATGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

BASE COUNT 54 a 45 c 29 g 85 t  
ORIGIN

Query Match 11.9%; Score 16; DB 35; Length 213;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GATTATCATAGACGA 64

Db 96 GATTATCATAGACGA 111

## RESULT 35

A1277304 262 bp mRNA EST 29-JAN-1999  
LOCUS qm67f07.x1 Soares\_placenta\_8to9weeks\_2Nbp8to9w Homo sapiens cDNA  
DEFINITION clone IMAGE:1893829 3', mRNA sequence.

ACCESSION A1277304.1 GI:3899572

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 262)

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Unpublished (1997)

On Jan 19, 1998 this sequence version replaced gi:2287250.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1206 Std Error: 0.00

Seq primer: -400p from Gibco

High quality sequence stop: 241.

Location/Qualifiers

1. .262  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone\_lib="Soares\_placenta\_8to9weeks\_2Nbp8to9w"

/dev\_stage="two placentae: one from 8 weeks and another  
from 9 weeks post conception"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: p77T3D (pharmacía) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TTTACCAATCTGAAGTGGAGCGCGCGGATTTTTTTTTTTT 3']

double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacía), digested with Not I and cloned into

the Not I and Eco RI sites of a modified p77T3 vector  
(Pharmacía). Library constructed by Bento Soares and

M.Fatima Bonaído."

BASE COUNT 74 a 58 c 37 g 93 t  
ORIGIN

Query Match 11.9%; Score 16; DB 44; Length 262;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GATTATCATAGACGA 64

Db 88 GATTATCATAGACGA 103

## RESULT 36

AQ663218 263 bp DNA GSS 23-JUN-1999  
LOCUS HS\_5463\_B2\_D01\_SP6E RPCI-11 Human Male BAC Library Homo sapiens  
DEFINITION genomic clone Plate=1039 Col=2 Row=H, genomic survey sequence.

ACCESSION AQ663218.1 GI:5170986

VERSION GSS.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 263)

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallaceu.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htsc.washington.edu

Plate: 1039 row: H column: 2

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 263.

Location/Qualifiers

1. .263  
/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Plate=1039 Col=2 Row=H"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Genomic sequence of BAC ends"

```

BASE COUNT      68 a      45 c      56 g      94 t
ORIGIN

Query Match      11.9%; Score 16; DB 105; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 CATAGAGCAATAATAA 71
|||||
Db 167 CATAGAGCAATAATAA 152

RESULT 37
T02730
LOCUS      T02730      272 bp      DNA      EST      29-NOV-1993
DEFINITION      0077M3 gmbpfH3.1, G. Roman Reddy Plasmodium falciparum genomic
clone PF0077M, mRNA sequence.
ACCESSION      T02730
VERSION      T02730
KEYWORDS
SOURCE
ORGANISM      malaria parasite P. falciparum.
REFERENCE
AUTHORS      Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C.
and Dame,J.B.
1 (bases 1 to 272)
TITLE      Gene sequence tags from Plasmodium falciparum genomic DNA fragments
prepared by the genease activity of mung bean nuclease
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
MEDLINE      94052193
COMMENT      Other ESTs: 0077M7
Contact: Dame JB
Department of Pathobiology
University of Florida
2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: dame@icbr.ifas.ufl.edu
Seq Primer: T3.
Location/Qualifiers
1. .272
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="PF0077M"
/clone_lib="gmbpfH3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran
HB3 isolate cultured in vitro, was digested with mung bean
nuclease in the presence of 30% formamide at 500C
(Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). The ends of the
fragments were polished using T4 DNA polymerase, and the
fragments were ligated to EcoR V-cleaved and
dephosphorylated pBluescript SK(+). Recombinant plasmids
transformed E. coli XL1-Blue."
74 a      33 c      38 g      111 t      16 others

BASE COUNT      74 a      33 c      38 g      111 t
ORIGIN

Query Match      11.9%; Score 16; DB 20; Length 272;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 37 TTGGTTATTAAATT 52
|||||
Db 55 TTGGTTATTAAATT 70

RESULT 38
C34963
LOCUS      C34963      300 bp      mRNA      EST      18-OCT-1999
DEFINITION      0077M3 gmbpfH3.1, G. Roman Reddy Plasmodium falciparum genomic
clone PF0077M, mRNA sequence.
ACCESSION      C34963
VERSION      C34963
KEYWORDS
SOURCE
ORGANISM      malaria parasite P. falciparum.
REFERENCE
AUTHORS      Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C.
and Dame,J.B.
1 (bases 1 to 272)
TITLE      Gene sequence tags from Plasmodium falciparum genomic DNA fragments
prepared by the genease activity of mung bean nuclease
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
MEDLINE      94052193
COMMENT      Other ESTs: 0077M7
Contact: Dame JB
Department of Pathobiology
University of Florida
2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: dame@icbr.ifas.ufl.edu
Seq Primer: T3.
Location/Qualifiers
1. .272
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="PF0077M"
/clone_lib="gmbpfH3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran
HB3 isolate cultured in vitro, was digested with mung bean
nuclease in the presence of 30% formamide at 500C
(Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). The ends of the
fragments were polished using T4 DNA polymerase, and the
fragments were ligated to EcoR V-cleaved and
dephosphorylated pBluescript SK(+). Recombinant plasmids
transformed E. coli XL1-Blue."
74 a      33 c      38 g      111 t      16 others

BASE COUNT      74 a      33 c      38 g      111 t
ORIGIN

```

```

DEFINITION      C34963 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
Caenorhabditis elegans cDNA clone yk418g1 3', mRNA sequence.
ACCESSION      C34963
VERSION      C34963.1
KEYWORDS
SOURCE
ORGANISM      Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 300)
AUTHORS      Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE      Expression map of the C.elegans genome
JOURNAL      Unpublished (1996)
COMMENT      On May 8, 1995 this sequence version replaced gi:801504.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_lib="yk418g1"
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
84 a      65 c      58 g      90 t      3 others

BASE COUNT      84 a      65 c      58 g      90 t
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 ATTAGACAAATAAACC 114
|||||
Db 20 ATTAGACAAATAAACC 35

RESULT 39
C36512
LOCUS      C36512      300 bp      mRNA      EST      18-OCT-1999
DEFINITION      0077M3 gmbpfH3.1, G. Roman Reddy Plasmodium falciparum genomic
clone PF0077M, mRNA sequence.
ACCESSION      C36512
VERSION      C36512
KEYWORDS
SOURCE
ORGANISM      malaria parasite P. falciparum.
REFERENCE
AUTHORS      Reddy,G.R., Chakrabarti,D., Schuster,S.M., Ferl,R.J., Almira,E.C.
and Dame,J.B.
1 (bases 1 to 272)
TITLE      Gene sequence tags from Plasmodium falciparum genomic DNA fragments
prepared by the genease activity of mung bean nuclease
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 90, 9867-9871 (1993)
MEDLINE      94052193
COMMENT      Other ESTs: 0077M7
Contact: Dame JB
Department of Pathobiology
University of Florida
2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: dame@icbr.ifas.ufl.edu
Seq Primer: T3.
Location/Qualifiers
1. .272
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="PF0077M"
/clone_lib="gmbpfH3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"
/note="Vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran
HB3 isolate cultured in vitro, was digested with mung bean
nuclease in the presence of 30% formamide at 500C
(Vernick, K.D., Imberiski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). The ends of the
fragments were polished using T4 DNA polymerase, and the
fragments were ligated to EcoR V-cleaved and
dephosphorylated pBluescript SK(+). Recombinant plasmids
transformed E. coli XL1-Blue."
74 a      33 c      38 g      111 t      16 others

BASE COUNT      74 a      33 c      38 g      111 t
ORIGIN

```

```

/strain="N2"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
BASE COUNT      84 a      62 c      61 g      91 t      2 others
ORIGIN
|||||

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
Db 30 ATTAGACAAATAAAC 45

RESULT 40
C51747 300 bp mRNA EST 11-SEP-1997
LOCUS C51747 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone_yk214b4 3', mRNA sequence.
ACCESSION C51747
VERSION C51747.1 GI:2389504
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1400939.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
Source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      76 a      64 c      59 g      93 t      8 others
ORIGIN
|||||

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
Db 25 ATTAGACAAATAAAC 40

RESULT 42
C52912 300 bp mRNA EST 11-SEP-1997
LOCUS C52912 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone_yk286b3 3', mRNA sequence.
ACCESSION C52912
VERSION C52912.1 GI:2390669
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
Source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      84 a      63 c      62 g      89 t      2 others
ORIGIN
|||||

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
Db 29 ATTAGACAAATAAAC 44

RESULT 41
C52581 300 bp mRNA EST 11-SEP-1997
LOCUS C52581 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone_yk260g2 3', mRNA sequence.

```

```

C52581
C52581.1 GI:2390338
EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1407336.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
Source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      76 a      64 c      59 g      93 t      8 others
ORIGIN
|||||

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAAATAAAC 114
Db 25 ATTAGACAAATAAAC 40

RESULT 42
C52912 300 bp mRNA EST 11-SEP-1997
LOCUS C52912 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone_yk286b3 3', mRNA sequence.
ACCESSION C52912
VERSION C52912.1 GI:2390669
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
FEATURES
Source
1..300
/organism="Caenorhabditis elegans"
/strain="CBI489 him-8(el489)"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"

```

```

/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT      84 a 63 c 62 g 89 t 2 others
ORIGIN

Query Match      11.9%; Score 16; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAATAAAC 114
|||||
Db 29 ATTAGACAATAAAC 44

RESULT 43
AV178626      300 bp mRNA      21-JUL-1999
DEFINITION    embryo Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
ACCESSION     AV178626
VERSION       AV178626
KEYWORDS      EST.
SOURCE        Caenorhabditis elegans.
ORGANISM      Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
              Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE     1 (bases 1 to 300)
AUTHORS       Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
              Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
              Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
              Nomoto,H.
              Expressed genes in C.elegans
              Unpublished (1999)
              On May 18, 1998 this sequence version replaced gi:3137286.
              Contact: Yuji Kohara
              Gene Library Lab
              National Institute of Genetics
              Yata 1111, Mishima, Shizuoka 411, Japan
              Tel: 81-559-81-6854
              Fax: 81-559-81-6855
              Email: ykohara@lab.nig.ac.jp.
              Location/Qualifiers
                1..300
                /organism="Caenorhabditis elegans"
                /strain="N2"
                /db_xref="taxon:6239"
                /clone="yk569d5"
                /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                hermaphrodite embryo"
                /sex="hermaphrodite"
                /dev_stage="embryo"
BASE COUNT      84 a 63 c 64 g 89 t
ORIGIN

Query Match      11.9%; Score 16; DB 62; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAATAAAC 114
|||||
Db 28 ATTAGACAATAAAC 43

RESULT 44
AV184617      300 bp mRNA      21-JUL-1999
LOCUS         AV184617 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION    embryo Caenorhabditis elegans cDNA clone yk657b11 3', mRNA
              sequence.

```

```

ACCESSION     AV184617      305 bp mRNA      11-NOV-1999
VERSION       AV3232351
KEYWORDS      EST.
SOURCE        Mus musculus cDNA clone 6330533L04 3', mRNA sequence.
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
              Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
              Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
              Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
              Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
              Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
              Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
              Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N.,
              Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
              Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
              RIKEN Mouse ESTs (Konno,H., et al.)
              Unpublished (1999)
              On Dec 20, 1995 this sequence version replaced gi:1133772.
              Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              Center,
              Tsukuba, Ibaraki 305-8565, Japan.
              E-mail: hayashizaki@life.tsukuba.ac.jp
              Location/Qualifiers
                1..305
                /organism="Mus musculus"
                /strain="N2"
                /db_xref="taxon:10116"
                /clone="yk657b11"
                /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                hermaphrodite embryo"
                /sex="hermaphrodite"
                /dev_stage="embryo"
BASE COUNT      84 a 65 c 62 g 89 t
ORIGIN

Query Match      11.9%; Score 16; DB 62; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 ATTAGACAATAAAC 114
|||||
Db 22 ATTAGACAATAAAC 37

RESULT 45
AV3232351/c
LOCUS         AV3232351 RIKEN full-length enriched, adult male medulla oblongata
DEFINITION    Mus musculus cDNA clone 6330533L04 3', mRNA sequence.
ACCESSION     AV3232351
VERSION       AV3232351
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
              Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
              Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I.,
              Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
              Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
              Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K.,
              Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y.,
              Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomimaga,N.,
              Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A.,
              Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
              RIKEN Mouse ESTs (Konno,H., et al.)
              Unpublished (1999)
              On Dec 20, 1995 this sequence version replaced gi:1133772.
              Contact: Yoshihide Hayashizaki
              Genome Exploration Research Group, Life Science Tsukuba Center,
              Center,
              Tsukuba, Ibaraki 305-8565, Japan.
              E-mail: hayashizaki@life.tsukuba.ac.jp
              Location/Qualifiers
                1..305
                /organism="Mus musculus"
                /strain="N2"
                /db_xref="taxon:10116"
                /clone="yk657b11"
                /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                hermaphrodite embryo"
                /sex="hermaphrodite"
                /dev_stage="embryo"
BASE COUNT      84 a 65 c 62 g 89 t
ORIGIN

```

Genome Science Laboratory  
The Institute of Physical and Chemical Research (RIKEN), Genomic  
Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: +81-298-36-9013  
Fax: +81-298-36-9098  
Email: genome-res@rtc.riken.go.jp/  
URL: http://genome.rtc.riken.go.jp/  
Sasaki, N., Izawa, N., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y.,  
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and  
Hayashizaki, Y.  
Transcriptional sequencing: A method for DNA sequencing using RNA  
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.

FEATURES  
source  
Location/Qualifiers  
1..305  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, adult male medulla  
oblongata"  
/sex="male"  
/tissue\_type="medulla oblongata"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site 1: Sali; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot-10.0 and subtraction to Rot-100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCGAGTTAATAAATATCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified phagescript KS(+) after bulk excision  
from Lambda FLC 1. Cloning sites, 5' end: Sali; 3' end:  
BamHI"

BASE COUNT 59 a 79 c 63 g 104 t  
ORIGIN  
Query Match 11.9%; Score 16; DB 72; Length 305;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 8 AATGAGTAAGAAGAAA 23  
|||||  
Db 21 AATGAGTAAGAAGAAA 6



GenCore version 4.5  
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 21:09:53 ; Search time 9481.22 Seconds

(without alignments)  
-63.090 Million cell updates/sec

Title: US-09-090-672B-10

Perfect score: 197

Sequence: 1 ATGGTACCAGTTTCAAATT.....AAGAATCTATTATGACT 197

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 821193 seqs, -1518192014 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_bal.\*  
2: gb\_ba2.\*  
3: gb\_cm.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pli.\*  
8: gb\_pl2.\*  
9: gb\_pri.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vi.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_in.\*  
21: em\_om.\*  
22: em\_or.\*  
23: em\_ov.\*  
24: em\_pat.\*  
25: em\_ph.\*  
26: em\_pl.\*  
27: em\_ro.\*  
28: em\_sts.\*  
29: em\_sy.\*  
30: em\_un.\*  
31: em\_vi.\*  
32: gb\_htg1.\*  
33: gb\_htg2.\*  
34: gb\_ini.\*  
35: gb\_in2.\*  
36: em\_bal.\*  
37: em\_ba2.\*  
38: em\_hum3.\*  
39: em\_hum4.\*  
40: gb\_pr4.\*  
41: gb\_htg3.\*  
42: gb\_htg4.\*  
43: gb\_htg5.\*

44: gb\_htg6.\*  
45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	197	100.0	177990	43	AC012514	Homo sapi
2	146	74.1	163288	45	AC012013	AC012514 Homo sapi
3	24	12.2	108056	8	ATAC005170	AC005170 Arabidops
4	23	11.7	161155	45	AC016910	AC016910 Homo sapi
5	21	10.7	167843	32	CNS01DSZ	AL122057 Homo sapi
6	21	10.7	170846	32	CNS01DTJ	AL132778 Homo sapi
7	20	10.2	55307	43	AC014684	AL132778 Homo sapi
8	20	10.2	72941	11	AC004227	AC014684 Drosophil
9	20	10.2	77340	43	AC012680	AC004227 Homo sapi
10	20	10.2	88071	11	AC002316	AC012680 Arabidops
11	20	10.2	162345	35	AC008415	AC002316 Homo sapi
12	19	9.6	5084	10	HS800872	AC008415 Drosophil
13	19	9.6	74673	40	AC005683	AL10218 Homo sapi
14	19	9.6	100368	11	HS514K20	AC005683 Homo sapi
15	19	9.6	111723	11	HSJ101K10	AL008731 Human DNA
16	19	9.6	116205	41	AC009457	AL080276 Human DNA
17	19	9.6	128247	33	AL133269	AC009457 Drosophil
18	19	9.6	135711	43	AC015596	AL133269 Homo sapi
19	19	9.6	152583	33	HSJ709L21	AC015596 Homo sapi
20	19	9.6	154129	41	AC009427	AL078646 Homo sapi
21	19	9.6	155164	11	AC005165	AC009427 Homo sapi
22	19	9.6	158454	42	AC011107	AC005165 Homo sapi
23	19	9.6	166925	33	AC007027	AC011107 Homo sapi
24	19	9.6	174050	44	AC009863	AC007027 Homo sapi
25	19	9.6	192823	43	AC012149	AC009863 Homo sapi
26	19	9.6	231949	32	HS1099D15	AC012149 Homo sapi
27	19	9.6	237523	1	RPXX04	AL035456 Homo sapi
28	19	9.1	1563	7	BNASLRGA	AJ235273 Rickettsi
29	18	9.1	1563	7	BNESLOCPB	L10737 Brassica na
30	18	9.1	1794	7	BNSL1R	Z21609 B.napus mRN
31	18	9.1	1825	12	MMU13839	Z26914 B.napus (We
32	18	9.1	2139	7	BOSLRG	U13839 Mus musculu
33	18	9.1	2787	34	PFSTARP	X52089 Brassica ge
34	18	9.1	2984	7	ATPCL15G	Z11508 A.thaliana
35	18	9.1	7186	2	AF102623	AF102623 Methanosa
36	18	9.1	10414	1	MX99740	X99740 Mycoplasma
37	18	9.1	26542	34	CEM18	Z68507 Caenorhabdi
38	18	9.1	38886	34	CELF46H5	U41543 Caenorhabdi
39	18	9.1	41402	41	AC008437	AC008437 Homo sapi
40	18	9.1	45617	44	AC016648	AC016648 Homo sapi
41	18	9.1	57784	42	AC012518	AC012518 Homo sapi
42	18	9.1	70390	41	AC010280	AC010280 Homo sapi
43	18	9.1	70952	7	AB024028	AB024028 Arabidops
44	18	9.1	71293	44	AC016217	AC016217 Homo sapi
45	18	9.1	88481	40	AF191071	AF191071 Homo sapi

# ALIGNMENTS

RESULT 1  
AC012514 AC012514 177990 bp DNA HTG 17-NOV-1999  
LOCUS Homo sapiens clone Rpl1-837P6, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 97  
DEFINITION unordered pieces.  
ACCESSION AC012514

AC012514.3 GI:6446645  
HTG: HTGS\_PHASE1.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;  
Eutheria: Primates: Catarrhini; Hominidae; Homo.  
1 (bases 1 to 177990)  
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,  
Bodotta,B., Bouck,B., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forecum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,  
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,  
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,  
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,  
Parish,B., Paxton,S., Payton,B., Perez,L., Fu,L.L., Quiles,M.,  
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,  
Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R., Tabor,P.,  
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,  
Weinstock,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J.,  
Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 177990)  
Worley,K.C.  
Direct Submission  
Submitted (29-OCT-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 17, 1999 this sequence version replaced gi:6249649.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 97 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 6897: contig of 6897 bp in length  
\* 6898 6917: gap of unknown length  
\* 6918 13470: contig of 6553 bp in length  
\* 13471 13490: gap of unknown length  
\* 13491 19070: contig of 5580 bp in length  
\* 19071 19090: gap of unknown length  
\* 19091 24387: contig of 5297 bp in length  
\* 24388 24407: gap of unknown length  
\* 24408 29534: contig of 5127 bp in length  
\* 29535 29554: gap of unknown length  
\* 29555 34620: contig of 5066 bp in length  
\* 34621 34640: gap of unknown length  
\* 34641 38332: contig of 3692 bp in length  
\* 38333 38352: gap of unknown length  
\* 38353 42003: contig of 3651 bp in length  
\* 42004 42023: gap of unknown length  
\* 42024 45245: contig of 3222 bp in length  
\* 45246 45265: gap of unknown length  
\* 45266 48418: contig of 3153 bp in length  
\* 48419 48438: gap of unknown length  
\* 48439 51515: contig of 3077 bp in length  
\* 51516 51535: gap of unknown length  
\* 51536 54451: contig of 2916 bp in length  
\* 54452 54471: gap of unknown length  
\* 54472 57383: contig of 2912 bp in length  
\* 57384 57403: gap of unknown length  
\* 57404 60180: contig of 2777 bp in length  
\* 60181 60200: gap of unknown length  
\* 60201 62927: contig of 2727 bp in length  
\* 62928 62947: gap of unknown length  
\* 62948 65653: contig of 2706 bp in length

* 134848	136178:	contig of 1331 bp in length	AC012013	163288 bp	DNA	HTG	09-DEC-1999
* 136179	136198:	gap of unknown length	LOCUS	Homo sapiens clone RP11-764L14, *** SEQUENCING IN PROGRESS ***, 14			
* 136199	137526:	contig of 1328 bp in length	DEFINITION	unordered pieces.			
* 137527	137546:	gap of unknown length	ACCESSION	AC012013			
* 137547	138868:	contig of 1322 bp in length	VERSION	AC012013.5 GI:6552766			
* 138869	138888:	gap of unknown length	KEYWORDS	HTG; HTGS_PHASE1.			
* 138889	140180:	contig of 1292 bp in length	SOURCE	human			
* 140181	140200:	gap of unknown length	ORGANISM	Homo sapiens			
* 140201	141400:	contig of 1200 bp in length	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
* 141401	141420:	gap of unknown length	Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
* 141421	142606:	contig of 1186 bp in length	1 (bases 1 to 163288)				
* 142607	142626:	gap of unknown length	Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,				
* 142627	143801:	contig of 1175 bp in length	Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,				
* 143802	143821:	gap of unknown length	Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,				
* 143822	144985:	contig of 1164 bp in length	David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,				
* 144986	145005:	gap of unknown length	Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,				
* 145006	146137:	contig of 1132 bp in length	Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,				
* 146138	146157:	gap of unknown length	Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,				
* 146158	147278:	contig of 1121 bp in length	Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,				
* 147279	147298:	gap of unknown length	Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,				
* 148407	148406:	contig of 1108 bp in length	Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,				
* 148427	148426:	gap of unknown length	Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,				
* 149533:	contig of 1107 bp in length		Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,				
* 149534	149553:	gap of unknown length	Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,				
* 149554	150650:	contig of 1097 bp in length	Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,				
* 150651	150670:	gap of unknown length	Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R.,				
* 150671	151766:	contig of 1076 bp in length	Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabbah,M.,				
* 151747	151766:	gap of unknown length	Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,				
* 151767	152839:	contig of 1073 bp in length	Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and				
* 152840	152859:	gap of unknown length	Gibbs,R.				
* 152860	153918:	contig of 1059 bp in length	Direct Submission				
* 153919	153938:	gap of unknown length	Unpublished				
* 153939	154996:	contig of 1058 bp in length	2 (bases 1 to 163288)				
* 154997	155016:	gap of unknown length	Worley,K.C.				
* 155017	156061:	contig of 1045 bp in length	Direct Submission				
* 156062	156081:	gap of unknown length	Submitted (19-OCT-1999) Human Genome Sequencing Center, Department				
* 156082	156081:	gap of unknown length	of Molecular and Human Genetics, Baylor College of Medicine, One				
* 157095	157114:	gap of unknown length	Baylor Plaza, Houston, TX 77030, USA				
* 157115	158125:	contig of 1011 bp in length	On Dec 10, 1999 this sequence version replaced gi:6478891.				
* 158126	158145:	gap of unknown length	----- Genome Center				
* 158146	159114:	contig of 969 bp in length	Center: Baylor College of Medicine				
* 159115	159134:	gap of unknown length	Center code: BCM				
* 159135	160066:	contig of 932 bp in length	Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a>				
* 160067	160086:	gap of unknown length	Contact: hgsc-help@bcm.tmc.edu				
* 160087	161008:	contig of 922 bp in length	----- Project Information				
* 161009	161028:	gap of unknown length	Center project name: HMKZ				
* 161029	161902:	contig of 874 bp in length	Center Clone name: RP11-764L14				
* 161903	161922:	gap of unknown length	----- Summary Statistics				
* 161923	162789:	contig of 867 bp in length	Sequencing vector: M13; L08821				
* 162790	162809:	gap of unknown length	Chemistry: Dye-terminator Big Dye; 100% of reads				
Query Match	100.0%;	Score 197; DB 43; Length 177990;	Assembly program: Phrap; version 0.980611				
Best Local Similarity	100.0%;	Pred. No. 9.5e-86;	Consensus quality: 150048 bases at least Q40				
Matches 197; Conservative	0; Mismatches	0; Indels	Consensus quality: 154828 bases at least Q30				
			Consensus quality: 157279 bases at least Q20				
			Estimated insert size: 154019; sum-of-contigs estimation				
			Quality coverage: 5.5x in Q20 bases; sum-of-contigs estimation				
QY 1	ATGGTACCAGTTTCAAAATTAACATGGTTATTTTACTGTGTGTCCTCCAAATTAACATTAG	60	-----				
Db 62031	AUGTACCCAGTTTCAAAATTAACATGGTTATTTTACTGTGTGTCCTCCAAATTAACATTAG	62090	* NOTE: This is a 'working draft' sequence. It currently				
			* consists of 14 contigs. The true order of the pieces				
QY 61	GGAATTTTGGTGTGGGCTCTATCTACTAGAAAAATATATATTTGGTGTCTGAAGATA	120	* is not known and their order in this sequence record is				
Db 62091	GGAATTTTGGTGTGGGCTCTGTATCTACTAGAAAAATATATATTTGGTGTCTGAAGATA	62150	* arbitrary. Gaps between the contigs are represented as				
			* runs of N, but the exact sizes of the gaps are unknown.				
QY 121	ATTTTGAATTAATACACAGACAGTTTAGCATTTCACAGACACAGTTGGCAGTTGAAG	180	* This record will be updated with the finished sequence.				
Db 62151	ATTTTGAATTAATACACAGACAGTTTAGCATTTCACAGACACAGTTGGCAGTTGAAG	62210	* as soon as it is available and the accession number will				
			* be preserved.				
QY 181	AATCTATTATATGACT	197	1 100825: contig of 100825 bp in length				
Db 62211	AATCTATTATATGACT	62227	* 100826 100845: gap of unknown length				
			* 100846 126300: contig of 25455 bp in length				
			* 126301 126320: gap of unknown length				
			* 126321 145983: contig of 19663 bp in length				
			* 145984 146003: gap of unknown length				
			* 146004 153267: contig of 7264 bp in length				
RESULT	2						

```

* 153268 153287: gap of unknown length
* 153288 154672: contig of 1385 bp in length
* 154673 154692: gap of unknown length
* 154693 155950: contig of 1258 bp in length
* 155951 155970: gap of unknown length
* 155971 157219: contig of 1249 bp in length
* 157220 157239: gap of unknown length
* 157240 158298: contig of 1059 bp in length
* 158299 158318: gap of unknown length
* 158319 159151: contig of 833 bp in length
* 159152 159171: gap of unknown length
* 159172 160001: contig of 830 bp in length
* 160002 160021: gap of unknown length
* 160022 160828: contig of 807 bp in length
* 160829 160848: gap of unknown length
* 160849 161652: contig of 804 bp in length
* 161653 162472: gap of unknown length
* 161673 162472: contig of 800 bp in length
* 162473 162493: gap of unknown length
* 162493 163288: contig of 796 bp in length.
FEATURES
    Location/Qualifiers
        1..163288
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="RP11-764L14"
BASE COUNT 47826 a 31744 c 33111 g 50313 t 294 others
ORIGIN

Query Match          74.1%; Score 146; DB 45; Length 163288;
Best Local Similarity 99.3%; Pred. No. 4.5e-61;
Matches 196; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTTACCCAGTTCAAAATTAACATGGTTATTTTACTTGTGTTCCCAAAATTTAAACATTAG 60
Db 122244 ATGTTACCCAGTTCAAAATTAACATGGTTATTTTACTTGTGTTCCCAAAATTTAAACATTAG 122185

QY 61 GGAATTTTGGTGTGGGCTGTATACACTAGAAAATATATATATTTGGTGTGCTGAAGATA 120
Db 122184 GGAATTTTGGTGTGGGCTGTATACACTAGAAAATATATATATTTGGTGTGCTGAAGATA 122125

QY 121 ATTTTGAGATAATTAGACAGAGAGTTTACATTTCAAGAACAAAGTTTGGCAGTTGAAG 180
Db 122124 ATTTTGAGATAATTAGACAGAGAGTTTACATTTCAAGAACAAAGTTTGGCAGTTGAAG 122065

QY 181 AATCTATTATATGACT 197
Db 122064 AATCTATTATATGACT 122048

RESULT 3
ATAC005170/c
LOCUS ATAC005170 108056 bp DNA PLN 13-OCT-1998
DEFINITION Arabidopsis thaliana chromosome II BAC T29E15 genomic sequence, complete sequence.
ACCESSION AC005170
VERSION AC005170.1 GI:3738313
KEYWORDS HTG
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE 1 (bases 1 to 108056)
AUTHORS Rounsley,S.D., Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Mason,T.M., Shen,M., Ronning,C.M., Fraser,C.M., Somerville,C.R. and Venter,J.C.
TITLE Arabidopsis thaliana chromosome II BAC T29E15 genomic sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108056)
AUTHORS Rounsley,S.D. and Lin,X.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1998) The Institute for Genomic Research, 9712

```

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org

3 (bases 1 to 108056)

Rounsley,S.D.

Direct Submission

Submitted (13-OCT-1998) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

On Oct 13, 1998 this sequence version replaced gi:3608124.

Address all correspondence to:

Steve Rounsley

The Institute for Genomic Research

9712 Medical Center Dr,

Rockville, MD 20850,

USA

e-mail: rounsley@tigr.org

BAC clone T29E15 is from Arabidopsis chromosome II and is contained

in the YAC clone C108C07.

The orientation of the sequence is from SP6 to T7 end of the BAC

clone.

Genes were identified by a combination of three methods: Gene

prediction programs including GRAIL (available by anonymous ftp

from arthur.epm.ornl.gov), Genefinder (Phil Green, University of

Washington), Genscan (Chris Burge,

http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene

(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the

complete sequence against a peptide database and the Arabidopsis

EST database at TIGR (http://www.tigr.org/tdb/at/at.html).

Annotated genes are named to indicate the level of evidence for

their annotation. Genes with similarity to other proteins are named

after the database hits. Genes without significant peptide

similarity but with EST similarity are named as 'unknown' proteins.

Genes without protein or EST similarity, that are predicted by more

than two gene prediction programs over most of their length are

annotated as 'hypothetical' proteins. Genes encoding tRNAs are

predicted by tRNAscan-SE (Sean Eddy,

http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are

identified by repeatmasker (Arian Smit,

http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of

genomic sequence that are not annotated as genes but have predicted

exons by GRAIL are annotated as misc features.

Location/Qualifiers

1..108056

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/chromosome="II"

/map="C108C07"

/clone="T29E15"

/complement(&lt;1..219)

/gene="T29E15.1"

/complement(&lt;1..219)

/gene="T29E15.1"

/note="F27L4.1"

complement(1..3525)

/note="overlap with BAC clone F27L4 (AC004482.1..3525)."

complement(&lt;1..210)

/gene="T29E15.1"

/note="unknown protein, 3' partial"

/codon\_start=1

/protein\_id="AAC63655.1"

/db\_xref="GI:3738314"

/translation="MFLILLVFCITVFVFWVTKGAGEAIEGKYKEYKLGDYSTMW

QKRVENGKNWKRISCLVESKVCSL"

270..445

/note="exon predicted by xgrail, quality

excellent\_shadowexon"

complement(688..704)

/note="exon predicted by xgrail, quality marginal"

complement(774..884)

/rpt\_family="(TAAA)n"

complement(777..848)

/rpt\_family="(TAAA)n"

1007..1054

/rpt\_family="AT-rich"

```
repeat_region      1103..1166
/rpt_family="AT-rich"
repeat_region      complement(1172..1819)
/rpt_family="(CAA)n"
repeat_region      1959..2114
/rpt_family="AT-rich"
repeat_region      2396..2422
/rpt_family="AT-rich"
/rpt_family="AT-rich"
misc_feature        complement(2741..2858)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
misc_feature        3114..3145
/note="exon predicted by xgrail, quality marginal"
repeat_region      3586..3691
/rpt_family="(CAAAA)n"
join(<3687..3956,4809..4933,5018..5055,5136..5246,
5359..5475,5694..5993)
/gene="T29E15.2"
<3687..5993
/gene="T29E15.2"
join(3687..3956,4809..4933,5018..5055,5136..5246,
5359..5475,5694..5764)
/gene="T29E15.2"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63656.1"
/db_xref="GI:3738315"
/translation="MKNVNVEFCSLSLVFLVPSFTSRSHIRFTYSAGASSPNRAIH
CMASDSQSGDGVSSPPNVAAYPSSSSSSASSALDFSLCTRLKTPPRAGWIKRDY
KDPESIAHRIKMLMALISSDIPGVNRDKMKMAIVHDIAEIVGDTIPSCGISKEE
NRRESIAEMHCKLLGGGRAKAEIAELWREYENSPEAKVYKDFDKVELILOALEY
BQGFQNICWAKASEIVSRRKQH"
complement(3861..3919)
/rpt_family="(GAA)n"
repeat_region      6423..6450
/rpt_family="AT-rich"
join(<6738..6789,6916..6986,7073..7154,7261..7376,
7475..>7603)
/gene="T29E15.3"
<6738..>7603
/gene="T29E15.3"
join(6738..6789,6916..6986,7073..7154,7261..7376,
7475..7603)
/gene="T29E15.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAC63657.1"
/db_xref="GI:3738316"
/translation="MSNNELLEIPMYLOFPPELKKMCSLYLTNKTDDNNVAFKVK
TRNNYCVRPNYGLILPKSTCKVLVTMOAKVEPDSMQSFEKFMIOQLASPGVTAK
VTREWFSGSHVVEENKIRVTVCSTTNTITSSPTRRGFFIQFC"
complement(7640..7728)
misc_feature        /note="exon predicted by xgrail, quality
excellent_shadowexon"
7643..7711
/note="exon predicted by xgrail, quality excellent"
7881..7943
/note="exon predicted by xgrail, quality marginal"
8342..8432
/note="exon predicted by xgrail, quality excellent"
repeat_region      complement(8539..8578)
/rpt_family="(TAAA)n"
8791..8854
/note="exon predicted by xgrail, quality marginal"
complement(8817..8852)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(8993..9066)
/misc_feature        /note="exon predicted by xgrail, quality
marginal_shadowexon"
9139..9226
/note="exon predicted by xgrail, quality marginal"
complement(9145..9242)
```

```
misc_feature        /note="exon predicted by xgrail, quality
marginal_shadowexon"
10777..10792
/note="exon predicted by xgrail, quality good"
complement(11093..11148)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
12104..12164
/note="exon predicted by xgrail, quality excellent"
complement(12228..12321)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
complement(12754..12788)
/note="exon predicted by xgrail, quality excellent"
12885..12931
/note="exon predicted by xgrail, quality good_shadowexon"
complement(12941..13003)
/note="exon predicted by xgrail, quality marginal"
13008..13124
/rpt_family="(TAAA)n"
13060..13125
/rpt_family="(CAAAA)n"
13258..13343
/rpt_family="(TAAA)n"
13429..13533
/rpt_family="AT-rich"
complement(13851..14008)
/rpt_family="AT-rich"
15830..15853
/note="exon predicted by xgrail, quality marginal"
15903..15991
/note="exon predicted by xgrail, quality excellent"
complement(16156..16274)
```

Query Match 12.24; Score 24; DB 8; Length 108056;

Best Local Similarity 100.08; Pred. No. 0.048;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 CAAGACAAGTTGGCAGTTGAAG 180

|||||

Db 20343 CAAGACAAGTTGGCAGTTGAAG 20320

RESULT 4

AC016910/c AC016910 161155 bp DNA HTG 08-DEC-1999  
LOCUS Homo sapiens clone RP11-460B1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 21  
DEFINITION unordered pieces.

ACCESSION AC016910

VERSION AC016910.1 GI:6539396

KEYWORDS HTG: HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 161155)

AUTHORS Waterston,R.H.

TITLE The sequence of Homo sapiens clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 161155)

AUTHORS Waterston,R.H.

TITLE Direct Submission

Submitted (08-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 21 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

```

* be preserved.
* 2185: contig of 2185 bp in length
* 2186: gap of unknown length
* 2204: contig of 2343 bp in length
* 4546: contig of 2343 bp in length
* 4547: gap of unknown length
* 4581: contig of 3397 bp in length
* 7979: gap of unknown length
* 7982: contig of 2079 bp in length
* 10059: contig of unknown length
* 10076: gap of unknown length
* 10077: contig of 2494 bp in length
* 12571: contig of unknown length
* 12588: gap of unknown length
* 12589: contig of 2492 bp in length
* 15081: gap of unknown length
* 15099: contig of 2585 bp in length
* 17684: gap of unknown length
* 20573: contig of 2892 bp in length
* 20574: gap of unknown length
* 20592: contig of 3108 bp in length
* 23699: contig of unknown length
* 23700: gap of unknown length
* 23717: gap of unknown length
* 23718: contig of 3983 bp in length
* 27701: gap of unknown length
* 27719: contig of 3640 bp in length
* 31359: gap of unknown length
* 36533: contig of 5157 bp in length
* 36534: gap of unknown length
* 40711: contig of 4160 bp in length
* 40712: gap of unknown length
* 40729: gap of unknown length
* 46976: contig of 6247 bp in length
* 46977: gap of unknown length
* 46994: gap of unknown length
* 54734: contig of 7740 bp in length
* 54735: gap of unknown length
* 54753: contig of 9535 bp in length
* 64288: gap of unknown length
* 64305: gap of unknown length
* 80245: contig of 15941 bp in length
* 80247: gap of unknown length
* 80265: contig of 13919 bp in length
* 94183: gap of unknown length
* 94201: gap of unknown length
* 94202: contig of 19101 bp in length
* 113303: gap of unknown length
* 113321: contig of 21224 bp in length
* 134545: gap of unknown length
* 134563: 161155: contig of 26593 bp in length.

FEATURES
  source
    Location/Qualifiers
      1..161155
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-4d0b1"

BASE COUNT 49973 a 31032 c 30999 g 48791 t 360 others
ORIGIN

Query Match 11.7%; Score 23; DB 45; Length 161155;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ATAATTTGAGATATAGACAA 140
      |||||
Db 151211 ATAATTTGAGATATAGACAA 151189

FEATURES
  source
    Location/Qualifiers
      1..161155
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-4d0b1"

BASE COUNT 49973 a 31032 c 30999 g 48791 t 360 others
ORIGIN

Query Match 11.7%; Score 23; DB 45; Length 161155;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 ATAATTTGAGATATAGACAA 140
      |||||
Db 151211 ATAATTTGAGATATAGACAA 151189

FEATURES
  source
    Location/Qualifiers
      1..161155
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-4d0b1"

BASE COUNT 49973 a 31032 c 30999 g 48791 t 360 others
ORIGIN

Query Match 11.7%; Score 21; DB 32; Length 167643;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ATAATTAGACAGACAGTTTA 149
      |||||
Db 108155 ATAATTAGACAGACAGTTTA 108175

FEATURES
  source
    Location/Qualifiers
      1..167643
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone.lib="RPC1-11"
        /clone="R-248EL4"
        /standard_name="RH99333"
        /note="matching EMBL:RH99333; Identified using the e-PCR software [G. Schuler]"

BASE COUNT 51600 a 30742 c 29577 g 50692 t 5032 others
ORIGIN

Genoscope.
Direct Submission
Submitted (20-OCT-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.
Contig order : 13 8 9 10 11 12, 1000 N's separate segments Contig
13 : length 52848 bp
Contig 8 : length 11004 bp
Contig 9 : length 24666 bp
Contig 10 : length 22438 bp
Contig 11 : length 32771 bp
Contig 12 : length 26081 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
  source
    Location/Qualifiers
      1..167643
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone.lib="RPC1-11"
        /clone="R-248EL4"
        /standard_name="RH99333"
        /note="matching EMBL:RH99333; Identified using the e-PCR software [G. Schuler]"

BASE COUNT 51600 a 30742 c 29577 g 50692 t 5032 others
ORIGIN

Query Match 10.7%; Score 21; DB 32; Length 167643;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ATAATTAGACAGACAGTTTA 149
      |||||
Db 108155 ATAATTAGACAGACAGTTTA 108175

FEATURES
  source
    Location/Qualifiers
      1..167643
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone.lib="RPC1-11"
        /clone="R-248EL4"
        /standard_name="RH99333"
        /note="matching EMBL:RH99333; Identified using the e-PCR software [G. Schuler]"

BASE COUNT 51600 a 30742 c 29577 g 50692 t 5032 others
ORIGIN

Genoscope.
Direct Submission
Submitted (20-OCT-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.
Contig order : 13 8 9 10 11 12, 1000 N's separate segments Contig
13 : length 52848 bp
Contig 8 : length 11004 bp
Contig 9 : length 24666 bp
Contig 10 : length 22438 bp
Contig 11 : length 32771 bp
Contig 12 : length 26081 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
  source
    Location/Qualifiers
      1..167643
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone.lib="RPC1-11"
        /clone="R-248EL4"
        /standard_name="RH99333"
        /note="matching EMBL:RH99333; Identified using the e-PCR software [G. Schuler]"

BASE COUNT 51600 a 30742 c 29577 g 50692 t 5032 others
ORIGIN

Query Match 10.7%; Score 21; DB 32; Length 167643;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ATAATTAGACAGACAGTTTA 149
      |||||
Db 108155 ATAATTAGACAGACAGTTTA 108175

FEATURES
  source
    Location/Qualifiers
      1..167643
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="14"
        /clone.lib="RPC1-11"
        /clone="R-248EL4"
        /standard_name="RH99333"
        /note="matching EMBL:RH99333; Identified using the e-PCR software [G. Schuler]"

BASE COUNT 51600 a 30742 c 29577 g 50692 t 5032 others
ORIGIN

Genoscope.
Direct Submission
Submitted (20-OCT-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.
Contig order : 25 21 20 26 30 29 23 27 22 28, 1000 N's separate
segments Contig 25 : length 4440 bp
Contig 21 : length 1785 bp

```

Contig 20 : length 1709 bp  
 Contig 26 : length 12108 bp  
 Contig 30 : length 32339 bp  
 Contig 29 : length 62473 bp  
 Contig 23 : length 2778 bp  
 Contig 27 : length 20623 bp  
 Contig 22 : length 1711 bp  
 Contig 28 : length 23487 bp  
 Contig 21 : length 1709 bp  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

source  
 Location/Qualifiers  
 1. .170846  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone\_lib="Caltech-D"  
 /clone\_fc="2568P8"  
 367. .484  
 /standard\_name="AA034525"  
 /note="matching EMBL:AA034525; Identified using the e-PCR  
 software (G. Schuler)"  
 12305. .12491  
 /standard\_name="AA599294"  
 /note="matching EMBL:AA599294; Identified using the e-PCR  
 software (G. Schuler)"

## STS

50617 a 30842 c 30677 g 49689 t 9021 others  
 BASE COUNT  
 ORIGIN

## Query Match

Best Local Similarity 10.7%; Score 21; DB 32; Length 170846;  
 Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ATAATTAGACACAGCTTTA 149

|||||  
 Db 169459 ATAATTAGACACAGCTTTA 169479

## RESULT 7

LOCUS AC014684/c  
 DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
 pieces.  
 AC014684  
 VERSION AC014684.1 GI:6436651  
 KEYWORDS HTG; HTGS\_PHASE2.  
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Phyloroides; Drosophilidae; Drosophila.  
 Direct Submission  
 Adams,M. and Venter,J.C.  
 Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA

COMMENT This sequence was identified as CDM:10210542 by the submitter.  
 For further information on this sequence e-mail to fly@celera.com.  
 \* NOTE: This is a 'working draft' sequence.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

## FEATURES

source  
 Location/Qualifiers  
 1. .55307  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 BASE COUNT 16942 a 11043 c 11010 g 16312 t  
 ORIGIN

## Query Match

10.2%; Score 20; DB 43; Length 55307;

Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 93 AAAATATATATATGGTGC 112  
 |||||  
 Db 26274 AAAATATATATATGGTGC 26255

## RESULT 8

LOCUS AC004227 72941 bp DNA PRI 26-FEB-1998  
 DEFINITION Homo sapiens chromosome 5, P1 clone 356a8 (LBNL H32), complete  
 sequence.  
 AC004227 L81383 L81382 L77059 L77058 L81381 L81629 L81804 AC001492  
 AC000982 AC000985 AC000983 AC001491 AC000984  
 VERSION AC004227.1 GI:2911718  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 72941)  
 AUTHORS Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,  
 Miguel,T., Lewis,K.D., Fridlyand,J., Alcivar,D., Benke,J.A.,  
 Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,  
 Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,  
 Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,  
 Abrajano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,  
 Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.  
 Sequencing of human chromosome 5

## TITLE

Unpublished

## JOURNAL

REFERENCE

2 (bases 1 to 72941)

Ricke,D.O. and Wagner,R.P.

Large Scale Sequence Analysis and Annotation with the Sequence

Comparison Analysis (SCAN) System

Unpublished

3 (bases 1 to 72941)

Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Subramanian,S.,

Miguel,T., Lewis,K.D., Fridlyand,J., Alcivar,D., Benke,J.A.,

Bondoc,M., Bowen,E., Chiang,A., Critz,P., Jaklevic,M.A., Lindo,K.,

Lindquist,K., Miller,C., Patel,S., Piscia,C., Riley,B.E.,

Rojeski,H., Sarmiento,R., Yu,C., Montenegro,M., Aerts,A., Chung,A.,

Abrajano,A., Baker,M., Gau,C., Jett,J., Ko,C., Beall,K.,

Woolley,J.P., Stultz,J.L., Kimmerly,W. and Martin,C.H.

Direct Submission

Submitted (26-FEB-1998) Human Genome Center, DOE Joint Genome

Institute, Lawrence Berkeley National Laboratory, MS 74-157,

Berkeley, CA 94720, U.S.A.

Sequence submitted by:

DOE Joint Genome Institute

For further information about this sequence, including its location

and relationship to other sequences, please visit the sequence

archive website (<http://www-hgc.lbl.gov/sequence-archive.html>) or

send email to [human@genome.lbl.gov](mailto:human@genome.lbl.gov)

FEATURES

Location/Qualifiers

1. .72941

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="5q"

/clone="356a8"

/chromosome="5"

/note="LBNL H32"

complement(1294. .1824)

/rpt\_family="Alu"

complement(2001. .2178)

/note="88% identity mouse V051C02.r1"

/db\_xref="dbEST:AA611267"

complement(<2020. .>50956)

/gene="K1AA0011B"

/note="K1AA0313-like protein"

complement(join(<2020. .2177,8946. .9186,18985. .19087,

```

22015. .22218.24986. .25093.27876. .28040.34059. .34211.
34797. .34955.37504. .37640.39747. .39924.50823. .>50956))
/gene="KIA001LB"
/codon_start=3
/evidence=not_experimental
/product="KIA001LB"
/protein_id="AAC04379.1"
/db_xref="GI:2911719"
/translation="LPADLTKMLHTENPHQVTHVSSQSGCSIASDSGSSLSDIYQ
ATESEVDVLTIRLEPGVDSEDEDEEIDRTDPLQGRDLVRECLKEKPAKTDGDD
VEQLRMHOLFPAFMGNSVRLCSVMIFVVEQAGALILEDGQELDSVTLINHT
VISHDPQGVENFAMTSGEGITPTLDKQVMHGIVTKVDDCOFCVIAOODVRLINHV
EKNTKVEEGEIVWVHEHRELDRTSGRKHIVIKATPERLMHLEHRSIVDTYIE
DLFTYRTFESPDVHGKLEWFKIDSUNDKVRIVLWVNNHNDPEGOPAMTRFL
BEFKNEDTMHNGHLLNIAACAARQVVOVQKASRESPLQFSLNGSEKGFIF
VEGPEKSAADSLGRGDQIMEVNGSFENTFMKAVEILRNTHLALTVKTNMFE
LLFTQEKGVPHPIAKKNSHNSIQHVPEDIEQTQSGKSKKVKANTVSGSRNK
IRKILDKTRFSILPPKLFSDGGLSQSDSDSIVGTRHCRHSLAIMPDPGLTSSSPDLL
OPTSMLDFFSNPS"
2580. .2637
/note="GRAIL 2 good exon, frame 1"
misc_feature
3601. .3903
/rpt_family="Alu"
repeat_region
5735. .5889
/rpt_family="Alu"
repeat_region
complement(5874. .6290)
/rpt_family="Alu"
repeat_region
6487. .6899
/note="97% identity yf1lb11.sl"
misc_feature
/db_xref="dbEST:R06850"
complement(7362. .7859)
/gene="KIA001LB"
/note="95% identity yf1lb11.rl"
misc_feature
/db_xref="dbEST:R0689"
complement(8946. .9164)
/gene="KIA001LB"
/note="GRAIL 2 good exon, frame 1"
misc_feature
complement(8959. .9186)
/gene="KIA001LB"
/note="61% identity KIAA0313"
repeat_region
complement(9570. .9793)
/rpt_family="MER8"
repeat_region
complement(9926. .10218)
/rpt_family="AlusX"
repeat_region
complement(10364. .10657)
/rpt_family="Alu"
repeat_region
10658. .10963
/rpt_family="Alu"
repeat_region
complement(11742. .12035)
/rpt_family="Alu"
repeat_region
12493. .12800
/rpt_family="Alu"
repeat_region
13239. .13547
/rpt_family="Alu"
repeat_region
13656. .13678
/note="(T)23"
/rpt_type=tandem
/rpt_unit=r
repeat_region
complement(13658. .14011)
/rpt_family="AlusC"
repeat_region
complement(14396. .14675)
/rpt_family="Alu"
repeat_region
complement(16140. .16255)
/rpt_family="Alu"
repeat_region
16160. .16180
/note="(T)21"
/rpt_type=tandem
/rpt_unit=r
repeat_region
complement(16458. .17179)
/rpt_family="LTR12"
repeat_region
complement(18979. .19084)
/gene="KIA001LB"
/note="GRAIL 2 excellent exon, frame 1"

```

```

misc_feature
complement(18985. .19086)
/gene="KIA001LB"
/note="82% identity KIAA0313"
complement(20614. .20759)
/rpt_family="MER20"
repeat_region
20960. .21305
/rpt_family="MER1B"
misc_feature
/gene="KIA001LB"
complement(22015. .22218)
/note="GRAIL 2 excellent exon, frame 2"
/gene="KIA001LB"
complement(22020. .22220)
/note="74% identity KIAA0313"
22476. .22511
/note="(ATT)9"
/rpt_type=tandem
/rpt_unit=ATT
repeat_region
complement(22488. .22823)
/rpt_family="Alu"
misc_feature
23112. .23397
/note="99% identity aa48dl2.sl"
/db_xref="dbEST:AA490888"
24394. .24751
/rpt_family="Alu"
repeat_region
complement(24986. .25093)
/gene="KIA001LB"
/note="GRAIL 2 good exon, frame 1"
complement(24995. .25096)
/gene="KIA001LB"
/note="94% identity KIAA0313"
25480. .25832
/rpt_family="Alu"
repeat_region
complement(25781. .27005)
/rpt_family="L1"
misc_feature
complement(27877. .28035)
/gene="KIA001LB"
/note="81% identity KIAA0313"
complement(28688. .28900)
/rpt_family="L1"
repeat_region
28703. .29059
/rpt_family="AlusP"
29015. .29036
/note="(A)22"
/rpt_type=tandem
/rpt_unit=A
repeat_region
complement(29325. .29653)
/rpt_family="Alu"
repeat_region
30392. .30799
/rpt_family="Alu"
repeat_region
31178. .32802
/rpt_family="MER42"
repeat_region
complement(32183. .32556)
/rpt_family="AlusX"
repeat_region
33615. .33964
/rpt_family="Alu"
misc_feature
complement(34058. .34216)
/gene="KIA001LB"
/note="88% identity KIAA0313"
complement(34059. .34211)
/gene="KIA001LB"
/note="GRAIL 2 excellent exon, frame 0"
complement(34797. .34955)
/gene="KIA001LB"
/note="GRAIL 2 excellent exon, frame 0"

```

Query Match 10.2%; Score 20; DB 11; Length 72941;  
Best Local similarity 100.0%; Pred.No. 4.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 GAAAAATATATATTGGTG 111  
|||||

Db 44348 GAAAAATATATATTGGTG 44367



<p>RESULT 9</p> <p>ACOL12680/c</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p>	<p>77340 bp DNA HTG 10-NOV-1999</p> <p>Arabidopsis thaliana chromosome I clone TAMU-Tl1111, *** SEQUENCING IN PROGRESS ***, 1 ordered pieces.</p> <p>ACOL12680</p> <p>ACOL12680.2 GI:6325535</p> <p>HTG; HTGS_PHASE2.</p> <p>Arabidopsis thaliana</p> <p>Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Magnoliophyta; euclidyales; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.</p> <p>1 (bases 1 to 77340)</p> <p>Lin.X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.</p> <p>Arabidopsis thaliana 'TAMU' BAC 'tl111' genomic sequence near marker 'agp64'</p> <p>Unpublished</p> <p>2 (bases 1 to 77340)</p> <p>Lin.X. and Kaul,S.</p> <p>Direct Submission</p> <p>Submitted (03-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA, Xinetia.org</p> <p>On Nov 10, 1999 this sequence version replaced gi:6223631.</p> <p>* NOTE: This is a 'working draft' sequence. It currently</p> <p>* consists of 1 contigs. Gaps between the contigs</p> <p>* are represented as runs of N. The order of the pieces</p> <p>* is believed to be correct as given, however the sizes</p> <p>* of the gaps between them are based on estimates that have</p> <p>* provided by the submittor.</p> <p>* This sequence will be replaced</p> <p>* by the finished sequence as soon as it is available and</p> <p>* the accession number will be preserved</p> <p>* 1 77340: contig of 77340 bp in length.</p>	<p>FEATURES</p> <p>Source</p> <p>1..77340</p> <p>/organism="Arabidopsis thaliana"</p> <p>/db_xref="taxon:3702"</p> <p>/chromosome="I"</p> <p>/clone="TAMU-Tl111"</p> <p>BASE COUNT 24501 a 14111 c 13706 g 24916 t 6 others</p> <p>ORIGIN</p>	<p>Query Match 10.2%; Score 20; DB 43; Length 77340;</p> <p>Best Local Similarity 100.0%; Pred. No. 4.3:</p> <p>Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p> <p>Oy 84 TATCACTAGAAAAATATA 103</p> <p>     </p> <p>Db 5574 TATCACTAGAAAAATATA 5555</p> <p>     </p> <p>RESULT 10</p> <p>AC002316</p> <p>LOCUS</p> <p>DEFINITION</p> <p>ACCESSION</p> <p>VERSION</p> <p>KEYWORDS</p> <p>SOURCE</p> <p>ORGANISM</p>	<p>88071 bp DNA PRI 08-AUG-1998</p> <p>Homo sapiens chromosome 17, clone HCIT11023, complete sequence.</p> <p>AC002316</p> <p>AC002316.1 GI:3406048</p> <p>HTG.</p> <p>Human.</p> <p>Homo sapiens</p> <p>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.</p> <p>1 (bases 1 to 88071)</p> <p>Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.</p> <p>Homo sapiens chromosome 17, clone HCIT11023</p> <p>Unpublished</p> <p>2 (bases 1 to 88071)</p>	<p>TITLE</p> <p>JOURNAL</p> <p>REFERENCE</p> <p>AUTHORS</p> <p>COMMENT</p> <p>LOCATION/Qualifiers</p>	<p>Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA</p> <p>Submitted (03-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA</p> <p>4 (bases 1 to 88071)</p> <p>Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J., Depayre,E., Devon,K., Dewar,K., DuRette,B., Etamadi,S., Ferreira,P., Forrest,C., Gage,D., Gardyna,S., Gensheimer,S., Geraiery,K., Gilmarin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L., Mackenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Stilwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and Zody,M.</p> <p>Direct Submission</p> <p>Submitted (05-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA</p> <p>5 (bases 1 to 88071)</p> <p>Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraiery,K., Grant,G., Hagos,B., Headford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Torruella-Miller,I., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.</p> <p>Direct Submission</p> <p>Submitted (08-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA</p> <p>On Aug 8, 1998 this sequence version replaced gi:2828784.</p> <p>All repeats were identified using RepeatMasker: Smit, A.F.A. &amp; Green, P. (1996-1997)</p> <p>http://ftp.genome.washington.edu/RM/RepeatMasker.html.</p> <p>Location/Qualifiers</p>
---	--	---	--	--	---	---

```
source      1. 88071
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="HC1111023"
            /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
            1-194)"
            /map="17"
            /chromosome="17"
            complement(236..380)
            /rpt_family="MIR"
            382..679
            /rpt_family="AluSg"
            680..701
            /rpt_family="(TAAA)n"
            complement(706..796)
            /rpt_family="LMB4"
            complement(953..1279)
            /rpt_family="AluSg"
            complement(1335..1523)
            /rpt_family="AluSg/x"
            1531..1762
            /rpt_family="AluJo"
            1766..2198
            /rpt_family="MER65A"
            complement(2223..2515)
            /rpt_family="AluSx"
            complement(2522..2784)
            /rpt_family="LIME"
            3202..3513
            /rpt_family="AluSg"
            4041..4339
            /rpt_family="AluSg"
            complement(5016..5226)
            /rpt_family="LMB4"
            5255..5612
            /rpt_family="MLT1A1"
            5639..5940
            /rpt_family="AluSx"
            7258..7555
            /rpt_family="AluSg"
            7558..7844
            /rpt_family="AluSx"
            7846..7874
            /rpt_family="(TAAA)n"
            8127..8358
            /rpt_family="L2"
            complement(9267..9304)
            /rpt_family="MIR"
            9305..9506
            /rpt_family="MIR"
            10123..10242
            /rpt_family="L2"
            complement(10389..10682)
            /rpt_family="AluSg"
            complement(11431..11741)
            /rpt_family="AluSx"
            12502..12822
            /rpt_family="AluSx"
            12823..12981
            /rpt_family="AluSg"
            complement(13045..13551)
            /rpt_family="L2"
            13552..13864
            /rpt_family="AluSx"
            complement(13865..14044)
            /rpt_family="L2"
            14505..14721
            /rpt_family="L1MC5"
            14722..15022
            /rpt_family="AluJb"
            15023..15043
            /rpt_family="L1MC5"
            15044..15350
```

```
repeat_region      /rpt_family="AluSx"
15351..15553
/rpt_family="L1MC5"
complement(15562..15831)
/rpt_family="AluJb"
15832..16035
/rpt_family="L1MC5"
complement(16350..16639)
/rpt_family="AluSx"
18544..18829
/rpt_family="AluYb8"
18830..18852
/rpt_family="(TAAA)n"
20480..20505
/rpt_family="(CCCCG)n"
20661..20703
/rpt_family="GC-rich"
20720..20824
/rpt_family="GC-rich"
complement(21825..21867)
/rpt_family="A-rich"
22587..22716
/rpt_family="AluJb"
complement(23096..23192)
/rpt_family="L2"
complement(23372..23677)
/rpt_family="AluSg"
complement(23687..23990)
/rpt_family="AluY"
24158..24457
/rpt_family="AluY"
25844..25958
/rpt_family="AluSx"
complement(26157..26234)
/rpt_family="L2"
27152..27453
/rpt_family="AluSg"
27624..27821
/rpt_family="MIR"
27860..28029
repeat_region
```

Query Match 10.2% Score 20; DB 11: Length 88071;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TAGAAAAATATATATATGG 109

Db 12121 TAGAAAAATATATATGG 12140

```
RESULT 11
AC006415/c AC006415 162345 bp DNA INV 09-JUL-1999
LOCUS Drosophila melanogaster, chromosome 2L, region 40B-40C, BAC clone
DEFINITION BACR2A22, complete sequence.
ACCESSION AC006415
VERSION AC006415.11 GI:5430741
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 162345)
AUTHORS Celiniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karta,K.,
Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Ling,H.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Preiffer,B., Poon,L., Richards,S., Sequeira,A., Sethi,H.,
Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M.,
Zieran,L.L. and Rubin,G.M.
```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Sequencing of Drosophila chromosome 2L, region 40B-40C
Unpublished (1998)
2 (bases 1 to 162345)
Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhorff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (19-JAN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jul 9, 1999 this sequence version replaced gi:5209360.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
FEATURES
source
1. .162345
/organism="Drosophila melanogaster"
/strain="v: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="40B-40C"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pACE3, 6)"
/clone="BAC BACR222A22 (D533)"
BASE COUNT 48827 a 32587 c 32005 g 48926 t
ORIGIN
10.2%; Score 20; DB 35; Length 162345;
Query Match 10.2%; Score 20; DB 35; Length 162345;
Best Local Similarity 100.0%; Pred. No. 3.9; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 93 AAAAATATATATTTGGTGC 112
|||||
Db 136676 AAAAATATATATTTGGTGC 136657
|||||
RESULT 12
HSM800872
LOCUS Homo sapiens mRNA; CDNA DKF2p434A163 (from clone DKF2p434A163).
DEFINITION Homo sapiens mRNA; CDNA DKF2p434A163 (from clone DKF2p434A163).
ACCESSION AL110218
VERSION AL110218.1 GI:5817150
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5084)
Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
Direct Submission
Submitted (13-AUG-1999) MIPS, Am Klopferstr. 18a D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); sequenced by AGOWA within the cDNA
sequencing consortium of the German Genome Project. This clone is
available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1. .5084
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p434A163"
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="testis"
/notes="corresponding STS: EMBL:HS09965"
1. .4964
/gene="DKF2p434A163"
CDS
<1. .4964
/note="strong similarity to C. elegans T27B1.1"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAB53678.1"
/db_xref="GI:5817151"
/translation="EDWDGKKVICNFKIQTSAVTCLOWPAEYLIIVFGLAEGKVRANT
KTKNSIYGTESIVVSLTNCSGKILSGVADGTIVRFFDDEGGESGQKLVNHP
PPYALAWATNSIVAAGCDKRRIVAYGEGHMLQTFDYSRDPQREFTTAVSSPGQSV
LGSYDLRYNKNFELTPRSIWEAKPKETINLYTITLAWKRDGSRCLVCGGEQFD
CCLRRSYKKNFELTYVPSQIVKNSGTRVWLKSHVGYEVEVKILGKERYLVAH
TSETLLGLDNTNRLSELAWOGSGGNEKYFFENENVMIFNAGELTIVYEGNNDTIGS
VTFEFNPHLISVRIKRCQRTEDNKKLALIDIKTIAIVDLIGYNTGTVSHSRV
DWLENETGHKLFRDKRLRLHLDIESCKTILNFCSTMQWPGSDVLVAQNRSL
CVYNIEAPERTVMTFIRGDIVIGLGGKTEVMVEGVTVAITIDEGIEFGTAID
DGNIRATAPLETLEMTPETEAMWKTLSKLAERQLHIAERCSFALGOVKARFLHE
WDECIATAEAKGHGPALEKLRYSYQWMDTOQERAGELQESOGDGLAALSLLKAGL
PAXAARLVLTRELLANTELVEHITAAIKGELYERAGULFEKIHQPKALECYRKGN
AFMKAVELARLAFPEVVKLEEAWDHVLVQKDDAAINHYIEARCSIKAEALGAR
QWKKAIIYILDQRNTASKYYPVLAQYASLOEYEAELYTKGDRTKDAIDHYTAG
RWEQAHKAMKMRPEDVSIVLYITQAEKMEKQGYREAEVLYVVOEPLAITMYKKH
KLVDMDIRLVKHPDLISDTHLHGKELEAGRLOEAEVHYLEAEQWPKATVNYRAS
GLWEEAVRVARTOGANAKHVAYLWAKSLGGEEAVRLNKLGLLEAAVDHADNCSF
EFAPFELSRALKHKTPEVHLKYAMFLEDEGKFEEAEAFIRACKPEAVLUMFVHNQDW
EAAQRAEADHPDSVAEVLVGQARGALEEKDFQKAEGLLRRAQRPGALNLYKEAGLV
SDALRIC KDYPSQLEALOEYERATKKGARGVEGVFAQRIQWEGVQAGYSRVAVCYL
KVRDSNGSLAEKCKWMAELSIKFLPPQRNMEVLAIVAPQLIGIKGHSAAELYNL
DLVKEADAFIEGEEWNKAKRVAKELDPYEDVDYDOHYEFKFLNKOGLVSLGVGVIA
ALDLYVEOGWDKCIETATKQNYKILHKVVALYATHLIREGSSAQLALVVOHGAN
PQNFNIYKRIFDMVSSPGTNCAEYHSHWADLRDLVFNLCENLVKSEANSAPAEFEK
TMLLIIHYIATRSAAQSVKQETVAARLSVSLRDTQLLPVDFKAFYEAAPAAVGM
NMAFIFLRFELDLTDAIEGTLGDSDFQDTDIPFVPLPAKHVPEAEEREVDW
VLTVMDQRLEQVLPDRGAYEASLVAASTGVRLPCLITGYTPILRNKTEFRKPGKA
ANKDNWNKFLMAIKTSHSPVCQDVLKFIQWCGLPSTSPSEQ"
polyA_signal 5001. .5006
polyA_site 5011
BASE COUNT 1401 a 1143 c 1434 g 1106 t
ORIGIN
Query Match 9.6%; Score 19; DB 10; Length 5084;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 152 ATTTACAAGAACAACTTGTG 170
|||||
Db 675 ATTTACAAGAACAACTTGTG 693
|||||
RESULT 13
AC005683
LOCUS AC005683 74673 bp DNA PRI 05-MAY-1999
DEFINITION Homo sapiens clone RG062N11, complete sequence.
ACCESSION AC005683
VERSION AC005683.2 GI:4753225
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 74673)
AUTHORS Waterston,R.H.
```

The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 74673)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (15-SEP-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 3 (bases 1 to 74673)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (05-MAY-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On May 5, 1999 this sequence version replaced gi:4204336.  
 Location/Qualifiers  
 1. .74673  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RG062N11"  
 /clone="RG062N11"  
 23089 a 14472 c 14794 g 22318 t  
 BASE COUNT  
 ORIGIN  
 Query Match 9.68; Score 19; DB 40; Length 74673;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 72 TTGTGGGTCTGTTATCACT 90  
 |||||  
 Db 6100 TTGTGGGTCTGTTATCACT 6118  
 |||||  
 RESULT 14  
 HS514K20 100368 bp DNA PRI 23-NOV-1999  
 LOCUS  
 DEFINITION  
 Human DNA sequence from PAC 514K20 on chromosome 6p22.3-24.3. EST,  
 CA repeats and STS.  
 AL008731  
 AL008731.1 GI:2769540  
 6522.3-24.3; repeat polymorphism.  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 100368)  
 Wild.A.  
 Direct Submission  
 Submitted (09-JAN-1998) Chromosome 6 Project Group  
 (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquires:  
 humquery@sanger.ac.uk  
 On Jan 13, 1998 this sequence version replaced gi:2598467.  
 IMPORTANT: This sequence is the entire insert of clone 514K20.  
 During sequence assembly data are compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variations annotated may not be found in the sequence submission  
 corresponding to the overlapping clone as we submit sequences with  
 only a small overlap as described above.  
 This sequence was generated from part of bacterial clone contigs of  
 human chromosome 6, constructed by the Sanger Centre chromosome 6  
 mapping group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6/  
 This sequence has been finished according to sequence map criteria  
 as follows. An attempt is made to resolve all sequencing problems,  
 such as compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 The true left end of clone 514K20 is at 1 in this sequence. The  
 true right end of clone 514K20 is at 100368.  
 514K20 is from the library RPC13 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong.  
 For further details see http://bacpac.med.buffalo.edu/.  
 Location/Qualifiers  
 1. .100368  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p22.3-24.3"  
 /clone="RP3-514K20"  
 /clone.lib="RPC1-3"  
 239. 534  
 repeat\_region  
 /note="AluJo repeat: matches 1. .294 of consensus"  
 1723. .2047  
 misc\_feature  
 /note="match: 267338 STS containing (CA) repeat"  
 1855. .1877  
 repeat\_region  
 /note="11 copies of CA 100% conserved; differs from  
 267338"  
 1882. .2174  
 repeat\_region  
 /note="AluJo repeat: matches 2. .301 of consensus"  
 3503. .3808  
 repeat\_region  
 /note="AluX repeat: matches 2. .301 of consensus"  
 4742. .5056  
 repeat\_region  
 /note="AluJo repeat: matches 1. .301 of consensus"  
 5276. .5897  
 repeat\_region  
 /note="LMB3 repeat: matches 291. .923 of consensus"  
 6167. .6299  
 repeat\_region  
 /note="LMB3 repeat: matches 767. .908 of consensus"  
 9707. .9975  
 repeat\_region  
 /note="MLT1B repeat: matches 390. .113 of consensus"  
 10357. .10552  
 repeat\_region  
 /note="MLT1C repeat: matches 466. .268 of consensus"  
 10571. .11046  
 repeat\_region  
 /note="MER4A2 repeat: matches 1. .505 of consensus"  
 11042. .11286  
 repeat\_region  
 /note="MLT1C repeat: matches 247. .1 of consensus"  
 12252. .12370  
 repeat\_region  
 /note="MLT1B repeat: matches 5. .120 of consensus"  
 14219. .14506  
 misc\_feature  
 /note="match: 216457 STS containing (CA) repeat"  
 14279. .14319  
 repeat\_region  
 /note="20 copies of CA 100 % conserved; differs from  
 216457; 96S359"  
 15184. .15235  
 repeat\_region  
 /note="2 copies of 26 mer 98 % conserved"  
 18638. .18929  
 repeat\_region  
 /note="AluX repeat: matches 302. .6 of consensus"  
 19749. .20135  
 repeat\_region  
 /note="MLT1C repeat: matches 23. .466 of consensus"  
 21690. .21761  
 repeat\_region  
 /note="MIR repeat: matches 186. .256 of consensus"  
 23026. .23327  
 repeat\_region  
 /note="AluX repeat: matches 1. .302 of consensus"  
 23328. .23478  
 repeat\_region  
 /note="AluJb repeat: matches 122. .273 of consensus;  
 incomplete repeat"  
 24403. .24786  
 repeat\_region  
 /note="LMD2 repeat: matches 469. .884 of consensus"  
 24779. .24937  
 repeat\_region  
 /note="MER42B repeat: matches 1149. .1300 of consensus"  
 26000. .26282  
 repeat\_region  
 /note="AluJb repeat: matches 293. .1 of consensus"  
 29619. .29921  
 repeat\_region  
 /note="AluY repeat: matches 2. .301 of consensus"  
 31380. .31670  
 repeat\_region  
 /note="AluSq repeat: matches 303. .1 of consensus"  
 31687. .31753  
 repeat\_region  
 /note="MIR2 repeat: matches 141. .75 of consensus"  
 31926. .32455  
 repeat\_region  
 /note="LIMA8 repeat: matches 1035. .514 of consensus"  
 32475. .32737  
 repeat\_region  
 /note="LJMB1 repeat: matches 566. .302 of consensus"  
 32859. .33179  
 repeat\_region  
 /note="AluJo repeat: matches 302. .1 of consensus"

repeat\_region 33831..33913 /note="MLTIC repeat: matches 89. .1 of consensus"  
repeat\_region 34014..34307 /note="AluY repeat: matches 1. .298 of consensus"  
repeat\_region 35042..35117 /note="MIR2 repeat: matches 146. .74 of consensus"  
repeat\_region 35186..35487 /note="AluSg repeat: matches 1. .303 of consensus"  
repeat\_region 36019..36447 /note="MLTIC repeat: matches 466. .1 of consensus"  
repeat\_region 38175..38477 /note="AluSx repeat: matches 302. .1 of consensus"  
repeat\_region 38520..38643 /note="MIR2 repeat: matches 145. .16 of consensus"  
repeat\_region 39015..39237 /note="L1ME1 repeat: matches 432. .661 of consensus"  
repeat\_region 39457..39631 /note="AluDo repeat: matches 125. .299 of consensus;  
incomplete repeat"  
repeat\_region 39650..39741 /note="L1MA2 repeat: matches 964. .1055 of consensus"  
repeat\_region 40095..40319 /note="L1PA14 repeat: matches 572. .896 of consensus"  
repeat\_region 41998..42078 /note="MIR repeat: matches 144. .56 of consensus"  
repeat\_region 42369..42595 /note="MER28 repeat: matches 1. .230 of consensus"  
repeat\_region 42654..42911 /note="AluSp repeat: matches 258. .1 of consensus;  
incomplete repeat"  
repeat\_region 42914..43132 /note="RIGGER2 repeat: matches 2472. .2708 of consensus"  
repeat\_region 43331..43631 /note="AluDo repeat: matches 1. .300 of consensus"  
repeat\_region 43633..43673 /note="MERSA repeat: matches 69. .109 of consensus"  
repeat\_region 43866..44166 /note="AluSx repeat: matches 1. .301 of consensus"  
repeat\_region 45157..45192 /note="18 copies of 2 mer 89 % conserved"  
repeat\_region 46610..46915 /note="AluSx repeat: matches 302. .1 of consensus"  
repeat\_region 47459..47740 /note="AluSx repeat: matches 15. .297 of consensus"  
repeat\_region 48084..48229 /note="MLT1A2 repeat: matches 1. .149 of consensus"  
repeat\_region 48230..48585 /note="MLT1A1 repeat: matches 1. .365 of consensus"  
repeat\_region 48586..48820 /note="MLT1B repeat: matches 154. .374 of consensus"  
repeat\_region 49766..49791 /note="13 copies of 2 mer 100 % conserved"  
repeat\_region 49951..49982 /note="16 copies of 2 mer 88 % conserved"  
repeat\_region 50250..50551 /note="AluSp repeat: matches 1. .303 of consensus"  
repeat\_region 51359..51922 /note="L1ME3A repeat: matches 630. .58 of consensus"  
repeat\_region 51805..51923 /note="L1MA1 repeat: matches 178. .57 of consensus"  
repeat\_region 52531..52831 /note="AluSp repeat: matches 3. .302 of consensus"  
repeat\_region 53351..53654 /note="AluSp repeat: matches 1. .303 of consensus"  
repeat\_region 54460..54843 /note="MSTD repeat: matches 394. .1 of consensus"  
repeat\_region 55614..55914 /note="AluSp repeat: matches 1. .303 of consensus"  
repeat\_region 58484..58593 /note="MSTC repeat: matches 1. .110 of consensus"  
repeat\_region 58597..58702 /note="MSTC repeat: matches 299. .405 of consensus"  
repeat\_region 59438..59525

repeat\_region /note="L1MB7 repeat: matches 298. .386 of consensus"  
61103..61222 /note="AluY repeat: matches 1. .124 of consensus;  
incomplete repeat"  
repeat\_region 61225..62555 /note="PFR5 repeat: matches 2438. .807 of consensus"  
repeat\_region 62556..62807 /note="AluY repeat: matches 44. .301 of consensus;  
incomplete repeat"  
repeat\_region 62868..63041 /note="MER5A repeat: matches 182. .2 of consensus"  
repeat\_region 64497..64784 /note="AluDb repeat: matches 297. .1 of consensus"  
repeat\_region 66089..66144 /note="MIR repeat: matches 86. .141 of consensus"  
repeat\_region 67000..67303  
Query Match 9.6%; Score 19; DB 11; Length 100368;  
Best Local Similarity 100.0%; Pred.No.13;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 128 GATAATTAGACAAGACAGT 146  
|||||  
Db 82127 GATAATTAGACAAGACAGT 82145  
RESULT 15  
HSJ101K10/C  
LOCUS HSJ101K10 111723 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 101K10 on chromosome 6q25-26,  
complete sequence.  
ACCESSION AL080276  
VERSION AL080276.9 GI:5763753  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 111723)  
AUTHORS Phillimore,B.  
TITLE Direct Submission  
JOURNAL Submitted (28-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
COMMENT On Aug 24, 1999 this sequence version replaced gi:5679778.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence is  
the entire insert of clone 101K10. This sequence has been finished  
according to sequence map criteria as follows. An attempt is made  
to resolve all sequencing problems, such as compressions and  
repeats, but not necessarily within known annotated human repeat  
sequence elements (e.g. Alu). Where the sequence is ambiguous,  
there is an annotation using the 'unsure' feature key.  
This sequence was generated from part of bacterial clone contigs of  
human chromosome 6, constructed by the Sanger Centre Chromosome 6  
mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
101K10 is from the library RPCI-1 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.  
FEATURES  
Location/Qualifiers  
1..111723  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```

/chrosome="6"
/map="q25-26"
/clone_lib="RCCI-1"
/clone_RPI-101K10"

BASE COUNT      35439 a 20986 c 21291 g 34007 t
ORIGIN

Query Match
Best Local Similarity 100.08; Pred.No.12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 TTTTGAGATAATTAGACAA 140
      |||||
Db 22529 TTTTGAGATAATTAGACAA 22511

RESULT 16
AC009457/c
LOCUS
DEFINITION
AC009457 116205 bp DNA HTG 29-SEP-1999
Drosophila melanogaster chromosome 3 clone BACR25A06 (D1082)
RCCI-98 25.A.6 map 88E-88E strain y; cn bw sp, *** SEQUENCING IN
PROGRESS ***, 106 unordered pieces.
AC009457
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 116205)
Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chave, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shrir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 116205)
Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chave, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shrir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (23-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 29, 1999 this sequence version replaced gi:5931387.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 106 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 472: contig of 472 bp in length
* 473 552: gap of unknown length
* 553 1036: contig of 486 bp in length

```

```

1118: gap of unknown length
1178: contig of 630 bp in length
1828: gap of unknown length
2432: contig of 604 bp in length
2512: gap of unknown length
3254: contig of 742 bp in length
3334: gap of unknown length
4000: contig of 666 bp in length
4080: gap of unknown length
4818: contig of 738 bp in length
4898: gap of unknown length
5405: contig of 507 bp in length
5485: gap of unknown length
6563: contig of 1078 bp in length
6643: gap of unknown length
7712: contig of 1069 bp in length
7792: gap of unknown length
8542: contig of 750 bp in length
8632: gap of unknown length
9136: contig of 514 bp in length
9216: gap of unknown length
9811: contig of 595 bp in length
9851: gap of unknown length
10285: contig of 374 bp in length
10345: gap of unknown length
11017: contig of 672 bp in length
11097: gap of unknown length
11789: contig of 692 bp in length
11859: gap of unknown length
12484: contig of 615 bp in length
12584: gap of unknown length
13692: contig of 1128 bp in length
13772: gap of unknown length
14746: contig of 974 bp in length
14826: gap of unknown length
15789: contig of 963 bp in length
15869: gap of unknown length
16819: contig of 950 bp in length
16899: gap of unknown length
18094: contig of 1195 bp in length
18174: gap of unknown length
19338: contig of 1164 bp in length
19418: gap of unknown length
20443: contig of 1025 bp in length
20523: gap of unknown length
21557: contig of 1034 bp in length
21637: gap of unknown length
22483: contig of 846 bp in length
22563: gap of unknown length
23643: contig of 1080 bp in length
23723: gap of unknown length
24701: contig of 978 bp in length
24781: gap of unknown length
25914: contig of 1133 bp in length
25994: gap of unknown length
26977: contig of 983 bp in length
27057: gap of unknown length
28157: contig of 1100 bp in length
28237: gap of unknown length
29386: contig of 1149 bp in length
29466: gap of unknown length
30141: contig of 675 bp in length
30221: gap of unknown length
31046: contig of 825 bp in length
31126: gap of unknown length
32575: contig of 1449 bp in length
32655: gap of unknown length
33532: contig of 877 bp in length
33612: gap of unknown length
34152: contig of 540 bp in length
34232: gap of unknown length
35240: contig of 1008 bp in length
35320: gap of unknown length

```



```

RESULT 18
AC015596
LOCUS      AC015596      135711 bp      DNA      HTG      17-NOV-1999
DEFINITION Homo sapiens clone RP11-45A9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC015596
VERSION     AC015596.1 GI:6447091
KEYWORDS   HTG; HTGS_PHASE0; NULL.
SOURCE     human.

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 135711)
AUTHORS    Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
            Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
            Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
            Cooke, P., DeArctiano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
            Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
            Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
            Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
            Lechoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
            McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
            Morrow, J., Navlor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
            Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
            Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
            Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
            Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Project Information
            Center project name: L110
            Center clone name: 45_A_9
            -----
            * NOTE: This record contains 154 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            1      854: contig of 854 bp in length
            *      gap of unknown length
            *      855      1719: contig of 865 bp in length
            *      gap of unknown length
            *      1720      2573: contig of 854 bp in length
            *      gap of unknown length
            *      2574      3402: contig of 829 bp in length
            *      gap of unknown length
            *      3403      4268: contig of 866 bp in length
            *      gap of unknown length
            *      4269      5129: contig of 861 bp in length
            *      gap of unknown length
            *      5130      5994: contig of 865 bp in length
            *      gap of unknown length
            *      5995      6870: contig of 876 bp in length
            *      gap of unknown length
            *      6871      7748: contig of 878 bp in length
            *      gap of unknown length
            *      7749      8647: contig of 899 bp in length
            *      gap of unknown length
            *
            8648      9518: contig of 871 bp in length
            *      gap of unknown length
            *      9519      10383: contig of 865 bp in length
            *      gap of unknown length
            *      10384      11258: contig of 875 bp in length
            *      gap of unknown length
            *      11259      12107: contig of 849 bp in length
            *      gap of unknown length
            *      12108      12961: contig of 854 bp in length
            *      gap of unknown length
            *      12962      13808: contig of 847 bp in length
            *      gap of unknown length
            *      13809      14687: contig of 879 bp in length
            *      gap of unknown length
            *      14688      15544: contig of 857 bp in length
            *      gap of unknown length
            *      15545      16427: contig of 883 bp in length
            *      gap of unknown length
            *      16428      17287: contig of 860 bp in length
            *      gap of unknown length
            *      17288      18139: contig of 852 bp in length
            *      gap of unknown length
            *      18140      18995: contig of 856 bp in length
            *      gap of unknown length
            *      18996      19839: contig of 844 bp in length
            *      gap of unknown length
            *      19840      20691: contig of 852 bp in length
            *      gap of unknown length
            *      20692      21539: contig of 848 bp in length
            *      gap of unknown length
            *      21540      22401: contig of 861 bp in length
            *      gap of unknown length
            *      22401      23265: contig of 865 bp in length
            *      gap of unknown length
            *      23266      24137: contig of 872 bp in length
            *      gap of unknown length
            *      24138      25042: contig of 905 bp in length
            *      gap of unknown length
            *      25043      25939: contig of 897 bp in length
            *      gap of unknown length
            *      25940      26843: contig of 904 bp in length
            *      gap of unknown length
            *      26844      27743: contig of 900 bp in length
            *      gap of unknown length
            *      27744      28649: contig of 906 bp in length
            *      gap of unknown length
            *      28650      29533: contig of 884 bp in length
            *      gap of unknown length
            *      29534      30445: contig of 912 bp in length
            *      gap of unknown length
            *      30446      31316: contig of 871 bp in length
            *      gap of unknown length
            *      31317      32236: contig of 920 bp in length
            *      gap of unknown length
            *      32237      33110: contig of 874 bp in length
            *      gap of unknown length
            *      33111      33972: contig of 862 bp in length
            *      gap of unknown length
            *      33973      34832: contig of 860 bp in length
            *      gap of unknown length
            *      34833      35743: contig of 911 bp in length
            *      gap of unknown length
            *      35744      36575: contig of 832 bp in length
            *      gap of unknown length
            *      36576      37454: contig of 879 bp in length
            *      gap of unknown length
            *      37455      38318: contig of 864 bp in length
            *      gap of unknown length
            *      38319      39201: contig of 883 bp in length
            *      gap of unknown length
            *      39202      40048: contig of 847 bp in length
            *      gap of unknown length
            *      40049      40918: contig of 870 bp in length
            *      gap of unknown length

```



\* gap of unknown length  
\* 40919 41800: contig of 882 bp in length  
\* gap of unknown length  
\* 41801 42686: contig of 885 bp in length  
\* gap of unknown length  
\* 42687 43548: contig of 862 bp in length  
\* gap of unknown length  
\* 43549 44392: contig of 844 bp in length  
\* gap of unknown length  
\* 44393 45271: contig of 879 bp in length  
\* gap of unknown length  
\* 45272 46139: contig of 868 bp in length  
\* gap of unknown length  
\* 46140 47065: contig of 926 bp in length  
\* gap of unknown length  
\* 47066 47930: contig of 865 bp in length  
\* gap of unknown length  
\* 47931 48798: contig of 868 bp in length  
\* gap of unknown length  
\* 48799 49684: contig of 886 bp in length  
\* gap of unknown length  
\* 49685 50537: contig of 833 bp in length  
\* gap of unknown length  
\* 50538 51400: contig of 863 bp in length  
\* gap of unknown length  
\* 51401 52280: contig of 880 bp in length  
\* gap of unknown length  
\* 52281 53155: contig of 875 bp in length  
\* gap of unknown length  
\* 53156 53999: contig of 844 bp in length  
\* gap of unknown length  
\* 54000 54846: contig of 847 bp in length  
\* gap of unknown length  
\* 54847 55691: contig of 845 bp in length  
\* gap of unknown length  
\* 55692 56570: contig of 879 bp in length  
\* gap of unknown length  
\* 56571 57443: contig of 873 bp in length  
\* gap of unknown length  
\* 57444 58326: contig of 883 bp in length  
\* gap of unknown length  
\* 58327 59202: contig of 876 bp in length  
\* gap of unknown length  
\* 59203 60078: contig of 876 bp in length  
\* gap of unknown length  
\* 60079 60939: contig of 861 bp in length  
\* gap of unknown length  
\* 60940 61811: contig of 872 bp in length  
\* gap of unknown length  
\* 61812 62684: contig of 873 bp in length  
\* gap of unknown length  
\* 62685 63588: contig of 904 bp in length  
\* gap of unknown length  
\* 63589 64454: contig of 866 bp in length  
\* gap of unknown length  
\* 64455 65322: contig of 868 bp in length  
\* gap of unknown length  
\* 65323 66187: contig of 865 bp in length

Query Match 9.6% Score 19; DB 43; Length 135711;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 AAAAATATATATATTGGTG 111  
|||||  
Db 108911 AAAAATATATATATTGGTG 108929

RESULT 19  
HSJ709L21 DNA HTG 03-DEC-1999  
LOCUS  
DEFINITION Homo sapiens chromosome 1 clone RP4-709L21 map q42.13-43, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, in unordered pieces.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AL078646  
AL078646.20 GI:6523750  
HTG; HTGS\_PHASE1.  
human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 152583)

Donnelly, S.  
Direct Submission

Submitted (03-DEC-1999), Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Dec 3, 1999 this sequence version replaced gi:6491758.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in  
progress and the release of this data is based on the understanding  
that the sequence may change as work continues. The sequence may  
be contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments. Contig-ID: 00157 Length: 1039bp

Contig-ID: 00362 Length: 1040bp

Contig-ID: 00988 Length: 1966bp

Contig-ID: 01010 Length: 1191bp

Contig-ID: 01548 Length: 124660bp

Contig-ID: 01722 Length: 1576bp

Contig-ID: 02108 Length: 4631bp

Contig-ID: 02181 Length: 1796bp

Contig-ID: 02245 Length: 1385bp

Contig-ID: 02591 Length: 2189bp

Contig-ID: 02919 Length: 1067bp

Contig-ID: 03086 Length: 1043bp.

\* NOTE: This is a 'working draft' sequence.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

FEATURES

source

Location/Qualifiers

1..152583

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="q42.13-43"

/clone="RP4-709L21"

/clone\_lib="RPCI-4"

BASE COUNT 35861 a 31061 c 33040 g 43820 t 8801 others

ORIGIN

Query Match 9.6% Score 19; DB 33; Length 152583;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 TTTTGAGATAATTAGACAA 140

|||||

Db 6696 TTTTGAGATAATTAGACAA 6714

|||||

RESULT 20

AC009427

LOCUS

AC009427 154129 bp DNA HTG 22-AUG-1999

DEFINITION Homo sapiens clone 45\_M\_3, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 23

unordered pieces.

ACCESSION

AC009427

VERSION

AC009427.1 GI:5757602

KEYWORDS

HTG; HTGS\_PHASE1.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 154129)

AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE

Homo sapiens, clone 45\_M\_3

JOURNAL

Unpublished

2 (bases 1 to 154129)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
 Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,  
 Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
 Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
 Hagos,B., Headford,A., Horton,L., Howland,J.C., Jones,C., Rann,L.,  
 Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
 Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
 Melgrim,J., Mella,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
 Naylor,J., Niloff,M., O'Connor,I., O'Donnell,P., Pavlin,B.,  
 Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
 Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
 Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.  
 Direct Submission  
 Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker: Smit, A.F.A. &  
 Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 23 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 185: contig of 185 bp in length  
 gap of unknown length  
 186 2563: contig of 2378 bp in length  
 gap of unknown length  
 2564 5880: contig of 3317 bp in length  
 gap of unknown length  
 5981 8528: contig of 2648 bp in length  
 gap of unknown length  
 8529 10743: contig of 2215 bp in length  
 gap of unknown length  
 10744 13869: contig of 3126 bp in length  
 gap of unknown length  
 13870 16720: contig of 2851 bp in length  
 gap of unknown length  
 16721 21145: contig of 4425 bp in length  
 gap of unknown length  
 21146 25413: contig of 4268 bp in length  
 gap of unknown length  
 25414 31523: contig of 6110 bp in length  
 gap of unknown length  
 31524 37054: contig of 5531 bp in length  
 gap of unknown length  
 37055 41745: contig of 4691 bp in length  
 gap of unknown length  
 41746 47233: contig of 5488 bp in length  
 gap of unknown length  
 47234 55611: contig of 8378 bp in length  
 gap of unknown length  
 55612 63950: contig of 8339 bp in length  
 gap of unknown length  
 63951 71243: contig of 7293 bp in length  
 gap of unknown length  
 71244 79316: contig of 8073 bp in length  
 gap of unknown length  
 79317 88629: contig of 9313 bp in length  
 gap of unknown length  
 88630 99510: contig of 10881 bp in length  
 gap of unknown length  
 99511 115536: contig of 16026 bp in length  
 gap of unknown length  
 115537 128158: contig of 12622 bp in length  
 gap of unknown length  
 128159 149566: contig of 21408 bp in length

\* \* 149567 154129: contig of 4563 bp in length.  
 Location/Qualifiers  
 1. 154129  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="45 M.3"  
 /clone\_lib="RPCI-11 Human Male BAC"  
 BASE COUNT 40465 a 35517 c 34138 g 42930 t 1079 others  
 ORIGIN

Query Match 9.6% Score 19; DB 41; Length 154129;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 ATTTACAAGACAAAGTTTG 170  
 |||||  
 Db 110173 ATTTACAAGACAAAGTTTG 110191

RESULT 21  
 AC005165  
 LOCUS AC005165 155164 bp DNA PRI 20-JUN-1998  
 DEFINITION Homo sapiens BAC clone RG242H14 from 7p15-p21, complete sequence.  
 ACCESSION AC005165  
 VERSION AC005165.1 GI:3242748  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 155164)  
 Becker,M., Tin-Wollam,A. and Harrison,M.  
 The sequence of Homo sapiens BAC clone RG242H14  
 JOURNAL Unpublished (1998)  
 2 (bases 1 to 155164)  
 Waterston,R.  
 Direct Submission  
 Submitted (20-JUN-1998) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 SUBMITTED BY: WUGSC  
 Department of Genetics  
 Washington University  
 St. Louis MO 63108, USA  
 http://genome.wustl.edu/gsc  
 mailto:sapiens@watson.wustl.edu

NOTE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate  
 chemistry; an attempt was made to resolve all sequencing problems,  
 such as compressions and repeats; all regions were covered by  
 sequence from more than one subclone; and the assembly was  
 confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and  
 sequencing collaboration between the NHGRI Chromosome 7 Mapping  
 Project (Eric D. Green, Director), John D. McPherson in the  
 Department of Genetics (Washington University), and the Washington  
 University Genome Sequencing Center. For additional information  
 about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send  
 mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 Clone RG242H14 is from a release of the human BAC library

CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelobAC11  
Selection: chloramphenicol

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RG241I02, 200 bp overlap.  
Actual start of this clone is at base position 88412 of RG241I02;  
actual end is at 155164 of RG242H14.

## FEATURES

source	Location/Qualifiers	
	1..155164	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="7"	
	/map="7p15-p21"	
	/clone_lib="CITB-HS-A"	
	/clone="RG242H14"	
STS	195..495	
	/db_xref="GI:435320"	
repeat_region	267..383	
	/rpt_family="MIR"	
repeat_region	449..578	
	/rpt_family="MIR"	
repeat_region	1139..1194	
	/rpt_family="(TGAA)n"	
repeat_region	1754..1805	
	/rpt_family="MIR"	
repeat_region	1900..2827	
	/rpt_family="L2"	
STS	complement(2701..3012)	
	/db_xref="GI:1916422"	
repeat_region	4372..4672	
	/rpt_family="Alu"	
repeat_region	5769..5814	
	/rpt_family="L2"	
repeat_region	6744..6897	
	/rpt_family="MIR"	
repeat_region	8941..9194	
	/rpt_family="Alu"	
repeat_region	9828..10013	
	/rpt_family="MER1_type"	
repeat_region	10198..10511	
	/rpt_family="MER1_type"	
repeat_region	11312..11486	
	/rpt_family="MIR"	
repeat_region	11916..12524	
	/rpt_family="L1"	
repeat_region	16979..17281	
	/rpt_family="Alu"	
repeat_region	18519..18641	
	/rpt_family="MIR"	
repeat_region	18811..18986	
	/rpt_family="L2"	
repeat_region	21003..21264	
	/rpt_family="L1"	
repeat_region	24656..24824	
	/rpt_family="MaLR"	
repeat_region	24825..25120	
	/rpt_family="Alu"	
repeat_region	25178..25371	
	/rpt_family="MaLR"	
repeat_region	25406..25468	
	/rpt_family="MIR"	
repeat_region	26231..26303	
	/rpt_family="MIR"	
misc_feature	complement(26404..26808)	
	/note="ubiquitin/CEP52 pseudogene"	
misc_feature	complement(26727..26811)	
	/note="match to EST AA368309 (NID:92020628)"	
repeat_region	27675..28202	
	/rpt_family="Retroviral"	

repeat_region	28936..29075	
	/rpt_family="MER1_type"	
repeat_region	29086..29384	
	/rpt_family="L1"	
repeat_region	29399..30562	
	/rpt_family="MER1_type"	
repeat_region	30610..30827	
	/rpt_family="MIR"	
repeat_region	32223..32520	
	/rpt_family="Alu"	
repeat_region	32817..32872	
	/rpt_family="MER4-group"	
repeat_region	32943..33227	
	/rpt_family="Alu"	
repeat_region	33530..33618	
	/rpt_family="L2"	
repeat_region	36813..37651	
	/rpt_family="Retroviral"	
repeat_region	37747..39045	
	/rpt_family="L1"	
repeat_region	39041..40642	
	/rpt_family="L1"	
repeat_region	40870..41028	
	/rpt_family="L2"	
repeat_region	41054..41327	
	/rpt_family="Alu"	
repeat_region	41402..41617	
	/rpt_family="L2"	
repeat_region	42224..42607	
	/rpt_family="MaLR"	
repeat_region	44185..44314	
	/rpt_family="MIR"	
repeat_region	45622..46028	
	/rpt_family="L1"	
repeat_region	46296..46607	
	/rpt_family="Alu"	
repeat_region	46635..47175	
	/rpt_family="L1"	
repeat_region	47198..47305	
	/rpt_family="MaLR"	
repeat_region	47306..47927	
	/rpt_family="L1"	
repeat_region	47960..48063	
	/rpt_family="L1"	
repeat_region	48063..48649	
	/rpt_family="L1"	
repeat_region	49199..49240	
	/rpt_family="L2"	
repeat_region	49664..49739	
	/rpt_family="MIR"	
repeat_region	50037..50111	
	/rpt_family="MIR"	
repeat_region	50469..50569	
	/rpt_family="L2"	
repeat_region	50763..50814	
	/rpt_family="L2"	
repeat_region	50869..51031	
	/rpt_family="MIR"	
repeat_region	51061..51399	
	/rpt_family="L2"	
repeat_region	52484..52593	
	/rpt_family="MIR"	
repeat_region	53714..54005	
	/rpt_family="Alu"	
repeat_region	54029..54149	
	/rpt_family="Alu"	
repeat_region	54160..54360	
	/rpt_family="L2"	
repeat_region	55960..56143	
	/rpt_family="MER1_type"	
repeat_region	56687..56988	
	/rpt_family="Alu"	
repeat_region	57413..57511	

```

1. .100923
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

```

/clone="DJ0832014"
BASE COUNT 53634 a 29738 c 31389 g 52128 t 36 others
ORIGIN

Query Match 9.6%; Score 19; DB 33; Length 166925;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 ACTAGAAAAATATATAT 106
|||||
Db 69586 ACTAGAAAAATATATAT 69568

RESULT 24
AC009863/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-573D15 map 3, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
AC009863
VERSION AC009863.2 GI:6514025
KEYWORDS HTG; HGVS_PHASE1; HGVS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 174050)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 3, clone RP11-573D15
Unpublished
REFERENCE 2 (bases 1 to 174050)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,B., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,I., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (03-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 3, 1999 this sequence version replaced gi:5822695.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genom Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1835
Center clone name: 573_D_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 139143 bases at least Q40
Consensus quality: 158720 bases at least Q30
Consensus quality: 168324 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 174050; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3771 contig of 3770 bp in length
* gap of unknown length
* 6298: contig of 2528 bp in length
* gap of unknown length
* 6299: contig of 2776 bp in length
* gap of unknown length
* 9075: contig of 3452 bp in length
* gap of unknown length
* 12527: contig of 3319 bp in length
* gap of unknown length
* 15846: contig of 5649 bp in length
* gap of unknown length
* 21495: contig of 5506 bp in length
* gap of unknown length
* 27001: contig of 7431 bp in length
* gap of unknown length
* 34432: contig of 9530 bp in length
* gap of unknown length
* 43962: contig of 9569 bp in length
* gap of unknown length
* 53531: contig of 13245 bp in length
* gap of unknown length
* 66776: contig of 14057 bp in length
* gap of unknown length
* 80833: contig of 12184 bp in length
* gap of unknown length
* 93017: contig of 13527 bp in length
* gap of unknown length
* 106544: contig of 20923 bp in length
* gap of unknown length
* 127467: contig of 46584 bp in length.
* Location/Qualifiers
* 1..174050
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="3"
* /map="3"
* /clone="Rp11-573D15"
* /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 48307 a 38133 c 37894 g 49708 t 8 others
ORIGIN

Query Match 9.6%; Score 19; DB 44; Length 174050;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 TAGAAAAATATATATATG 108
|||||
Db 70487 TAGAAAAATATATATG 70469

RESULT 25
AC012149/c
LOCUS
DEFINITION Homo sapiens chromosome 3q27 clone RP11-119E3, *** SEQUENCING IN
PROGRESS ***, 29 unordered pieces.
AC012149
VERSION AC012149.1 GI:6091636
KEYWORDS HTG; HGVS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 192823)

```

```

AUTHORS      Muzoy,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
             Bodola,B., Boucks,J., Bowles,S., Brooks,A., Buhay,C., Bunac,C.,
             Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
             David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
             Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
             Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
             Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
             Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
             Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
             Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
             Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
             Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
             Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
             Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
             Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P.,
             Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wahbah,M., Watlington,S.,
             Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J.,
             Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 192823)
Worley,K.C.
Direct Submission
Submitted (21-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1      854: contig of 864 bp in length
*      865      1826: contig of 962 bp in length
*      1827      3148: contig of 1322 bp in length
*      3149      3091: contig of 843 bp in length
*      3092      4821: contig of 830 bp in length
*      4822      5883: contig of 1062 bp in length
*      5884      6754: contig of 871 bp in length
*      6755      7893: contig of 1139 bp in length
*      7894      8736: contig of 843 bp in length
*      8737      9534: contig of 798 bp in length
*      9535      10543: contig of 1009 bp in length
*      10544      11959: contig of 1416 bp in length
*      11960      12783: contig of 824 bp in length
*      12784      13565: contig of 782 bp in length
*      13566      14382: contig of 817 bp in length
*      14383      15924: contig of 1542 bp in length
*      15925      17543: contig of 1619 bp in length
*      17544      17976: contig of 433 bp in length
*      17977      20336: contig of 2360 bp in length
*      20337      26022: contig of 5686 bp in length
*      26023      32770: contig of 6748 bp in length
*      32771      36814: contig of 4044 bp in length
*      36815      45401: contig of 8587 bp in length
*      45402      56484: contig of 11083 bp in length
*      56485      77905: contig of 21421 bp in length
*      77906      100224: contig of 22319 bp in length
*      100225      127870: contig of 27846 bp in length
*      127871      157857: contig of 29987 bp in length
*      157858      192823: contig of 34966 bp in length.
FEATURES      Location/Qualifiers
             1..192823
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="3q27"
             /clone="RP11-119E3"
BASE COUNT   53592 a 42230 c 42416 g 54504 t      81 others
ORIGIN

```

```

Query Match          9.6%: Score 19; DB 43; Length 192823;
Best Local Similarity 100.0%: Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TAGAAAAATATATATATTG 108
   |||||
Db 106574 TAGAAAAATATATATATTG 106556

RESULT 26
HS1099D15/c
LOCUS   HS1099D15 231949 bp DNA HTG 23-NOV-1999
DEFINITION Homo sapiens chromosome 20 clone RP5-1099D15, *** SEQUENCING IN
           PROGRESS ***, in unordered pieces.
ACCESSION AL035456
VERSION    AL035456.23 GI:6018445
KEYWORDS   HTG; HTGS_P1ASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 231949)
AUTHORS    Blakey,S.
JOURNAL    Direct Submission
           Submitted (08-OCT-1999) Wellcome Trust Genome Campus, Hinxton,
           Cambridgeshire, CB10 1SA, UK. E-mail enquires:
           humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
           On Oct 11, 1999 this sequence version replaced gi:6002127.
COMMENT     IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: dj1099D15 Contig_ID: 00002 acc=AL035456
            Length: 1092 bp Unfinished: dj1099D15 Contig_ID: 00040
            acc=AL035456 Length: 1301 bp Unfinished: dj1099D15 Contig_ID:
            00147 acc=AL035456 Length: 4081 bp Unfinished: dj1099D15
            Contig_ID: 00229 acc=AL035456 Length: 1803 bp Unfinished:
            dj1099D15 Contig_ID: 00351 acc=AL035456 Length: 2366 bp
            Unfinished: dj1099D15 Contig_ID: 00365 acc=AL035456 Length: 1251
            bp Unfinished: dj1099D15 Contig_ID: 00461 acc=AL035456 Length:
            1369 bp Unfinished: dj1099D15 Contig_ID: 00522 acc=AL035456
            Length: 1508 bp Unfinished: dj1099D15 Contig_ID: 00588
            acc=AL035456 Length: 1153 bp Unfinished: dj1099D15 Contig_ID:
            00755 acc=AL035456 Length: 2228 bp Unfinished: dj1099D15
            Contig_ID: 00777 acc=AL035456 Length: 125952 bp Unfinished:
            dj1099D15 Contig_ID: 00939 acc=AL035456 Length: 1071 bp
            Unfinished: dj1099D15 Contig_ID: 01003 acc=AL035456 Length: 2707
            bp Unfinished: dj1099D15 Contig_ID: 01006 acc=AL035456 Length:
            1815 bp Unfinished: dj1099D15 Contig_ID: 01070 acc=AL035456
            Length: 1492 bp Unfinished: dj1099D15 Contig_ID: 01141
            acc=AL035456 Length: 1218 bp Unfinished: dj1099D15 Contig_ID:
            01155 acc=AL035456 Length: 1986 bp Unfinished: dj1099D15
            Contig_ID: 01161 acc=AL035456 Length: 1469 bp Unfinished:
            dj1099D15 Contig_ID: 01238 acc=AL035456 Length: 1845 bp
            Unfinished: dj1099D15 Contig_ID: 01487 acc=AL035456 Length: 1366
            bp Unfinished: dj1099D15 Contig_ID: 01496 acc=AL035456 Length:
            1588 bp Unfinished: dj1099D15 Contig_ID: 01886 acc=AL035456
            Length: 1330 bp Unfinished: dj1099D15 Contig_ID: 02082
            acc=AL035456 Length: 2237 bp Unfinished: dj1099D15 Contig_ID:
            02171 acc=AL035456 Length: 1520 bp Unfinished: dj1099D15
            Contig_ID: 02387 acc=AL035456 Length: 1793 bp Unfinished:
            dj1099D15 Contig_ID: 02398 acc=AL035456 Length: 2266 bp
            Unfinished: dj1099D15 Contig_ID: 02481 acc=AL035456 Length: 2668
            bp Unfinished: dj1099D15 Contig_ID: 02511 acc=AL035456 Length:
            1406 bp Unfinished: dj1099D15 Contig_ID: 02582 acc=AL035456
            Length: 2212 bp Unfinished: dj1099D15 Contig_ID: 02801
            acc=AL035456 Length: 2618 bp Unfinished: dj1099D15 Contig_ID:
            02962 acc=AL035456 Length: 4740 bp Unfinished: dj1099D15
            Contig_ID: 02963 acc=AL035456 Length: 1924 bp Unfinished:
            dj1099D15 Contig_ID: 03015 acc=AL035456 Length: 3036 bp
            Unfinished: dj1099D15 Contig_ID: 03046 acc=AL035456 Length: 1505

```

bp Unfinished: dJ1099D15 Contig\_ID: 03081 acc=AL035456 Length: 1537 bp Unfinished: dJ1099D15 Contig\_ID: 03084 acc=AL035456 Length: 1084 bp Unfinished: dJ1099D15 Contig\_ID: 03243 acc=AL035456 Length: 7634 bp Unfinished: dJ1099D15 Contig\_ID: 03255 acc=AL035456 Length: 2178 bp.  
\* NOTE: This is a "working draft" sequence.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

## FEATURES

Location/Qualifiers

1..231949

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="20"

/clone="RP3-1099D15"

/clone\_lib="RPC1-5"

BASE COUNT 58430 a 42805 c 43044 g 58039 t 29631 others

## ORIGIN

Query Match 9.6%; Score 19; DB 32; Length 231949;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 127 AGATAATTAGACAAGACAG 145

|||||

Db 35078 AGATAATTAGACAAGACAG 35060

## RESULT 27

RPXX04/c

LOCUS

DEFINITION

Rickettsia prowazekii strain Madrid E, complete genome; segment 4/4.

ACCESSION

AJ235273

VERSION

AJ235273.1

KEYWORDS

complete genome.

SOURCE

Rickettsia prowazekii.

ORGANISM

Bacteria; Proteobacteria; alpha subdivision; Rickettsiales; Rickettsiaceae; Rickettsiae; Rickettsia.

REFERENCE

1 (bases 1 to 237523)

AUTHORS

Andersson, S.G., Zomorodipour, A., Andersson, J.O.,

Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K.,

Eriksson, A.S., Winkler, H.H. and Kurland, C.G.

The genome sequence of Rickettsia prowazekii and the origin of mitochondria

NATURE 396 (6707), 133-140 (1998)

MEDLINE

99039499

REFERENCE

2 (bases 1 to 237523)

AUTHORS

Andersson, S.G.E.

TITLE

Direct Submission

JOURNAL

Submitted (11-NOV-1998) S.G.E. Andersson,

Siv. Anderssonsmobio.uu.se, Dept. of Molecular Biology, University

of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN

Location/Qualifiers

1..237523

/organism="Rickettsia prowazekii"

/strain="Madrid E"

/db\_xref="taxon:782"

164..1123

/gene="RP701"

164..1123

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="CAA15137.1"

/db\_xref="GI:3861238"

/translation="MAEPIFTLLFGKDRILLRNNOKALNKLHPIIPLOVNEKSHIS

DKERIFSELEKSYKECKIPINFIQISNNNNKPIFRAODLINAQNLNIKITIFEDRYN

EDQIYGLKLEIHEILKDPNNAENFKYYSQISRQFKDLDKDKQKHNDSNFNTQKVI  
NDSNKDQVEHITSYVDREKQSKPOGFLTRIFNFKIVFKQ"

gene

CDS

/gene="Rp702"

2074..3348

/gene="Rp702"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="CAA15138.1"

/db\_xref="GI:3861239"

/translation="MTTILVILIIIMLSALFAATETAITASSPGKIHKLKIAKNKR  
AKTVLEVLKKEKVGILTKGNSLINTLTATLFLISLGNMTIVASGVMAFLII  
VFAEYVPAIAVAPEOLITLKMSTIVILKLFKPINIALDYITKECFEIPFINLNPQ  
ISGTEEVKRVIERHQGGYKSERMIGLIDIRNMTVSEIMTHRSNIIAUNIDLPY  
ELIITKLSGAHTRIPLMQNRDNIIGLNKDLKALYENNNDDKKVDINTLLPFPW  
FIPONALVVDOLHAFERNHFAVCVDEYETGLGIIITLEDVIEEIVGPTTDEHRLNN  
ETIQSNTEFIKGTITIRIDINRELDNLSDEADANTAGLIITHKIARIPNOGEVIEIF  
NFKIILKKIAKIDSVKITVLPMTETISSE"

gene

CDS

/gene="Rp703"

3341..5353

/gene="Rp703"

/codon\_start=1

/transl\_table=11

/product="CYTOCHROME C-TYPE BIOGENESIS PROTEIN CCMF

(ccmf)"

/protein\_id="CAA15139.1"

/db\_xref="GI:3861240"

/translation="MSKIGNFLTLTNCLSGILTLIIIPWLDYLIKKHNNICLKFF  
PWICTSOVMTQALTFYASSCSFACFTLVYAFITSDFSVONFVLUHSLTAPLKI  
SGSSWSEHGLMFLQDLIVSCYIFCEDRMKLFESSIIISALQLESFPIYFTSN  
PHEMSFVKEGILNPMQDIALSIHPPLLYGVYSYAVPTIVICAFELIARIERLD  
NNLPGYFTGRTISDRISIRLLKIFACGGLFTTIGISLGSWAYRELNMGGLWF  
FDPVENISLFWLSGIMLHSHIVTTKTRMOWHTIISITFLVIFSTFLVRGCFI  
NSIHAFSSERWVLLVILCIIGIGLVYIIRLNCSEKTIYTFNNTNTHSNMNT  
TKEKALIGNIFLFLSILICATLYTIVALENDKPIIINEEFKFIIRVPIIC  
TIGLFTNSSIKKHIIIMLSLITFIILLIKFVSILTIWTSIFLMLHNVHLV  
KINFRSLKASSIMLGHFGVALIAPSTIMNSLQSEDFIGESGTSIHSLNNAKVT  
LONIKFAQKNYFOIAFEWLEDHNNITILKPENFYIEQKLSDESIIHSLLYDL  
YAVLSNIDGKIIHAKYIKPMGFIWIGVILTITAGFVIALIRKNSY"

gene

CDS

/gene="Rp704"

complement(7264..12195)

/gene="Rp704"

/codon\_start=1

/transl\_table=11

/product="CELL SURFACE ANTIGEN (sca5)"

/protein\_id="CAA15140.1"

/db\_xref="GI:3861241"

/translation="MAOKPNFLKKIISAGLVLTASTATIVAGFSVGAMGAOMYNRTTN  
AAATFDGIGFDOAGANIPVAPNSVITANANNPITENTPNHLSLELDOTLANDLAVT  
INEDTLGFTINIAOAKFFETVAAGKIILNITGOGITVQESNTINAGNALTKVHGG  
AALNANDLSGISTIFAAPSVLEFNLTPTQEPALTIGANSKIYNGNGTLNTING  
FIQVSDNIFAGIKIINDCCGLMFNTPDANTLNLOVGGNTINNGIDGKGLVLV  
SKNGAATFNVGTGLGNLKGIIIELTAAVAGKLIISQGAANAVITGDNGAGRAAGFI  
VSDNGNAATISGVYKAMVYIOSANAGQVFEHIVDVLGGTTFKFAVSKVITTE  
NSNFGSTFNGLDQIVVPDTKILKNFGIGDVKNNGTAGVITFNANGALVASADPN  
IAVTNINAEAGVAGVVELSGIHAEURLNGSGSIFKLADGTIVNGPNVONALMNNNA  
LAAGSILDGSAIITGIDINGGVAAQLHITTLANDASKI LALDGANI IGANVGAATF  
OANGGTIKLTNTONNVNFDLDTTDTKTVDVASSLTNNQTLTNGSTGVVANTKT  
LAQLNIGSKTILANGDVAIDELVIEKNGSVQLNHNHTLTKTINANGOGIIVAADP  
LNTNLTADGNILGSAENPLSTIHFATKAAANADSLNKGYNLYANNITLIDNANVGS  
LHFRSGGTISVSGTVGGQGHKLNLLDNGTIVFKLGDTFNGGTFKLGKSLQISLN  
NTTDTVESADNTGILEFNTPDTIVTLNKGAYFGVLKQVITSGPGNTVFNEIGNVG  
IVHGAANSISFENASLGTSLFSPGTPDLVLAIKSTVGNVTGDNFNAPIVVSGIDS  
MINNGOIIIGDKNIIALSLSGDSNTIVNANTLXSGRTKNGOVTLSGGMNPNPGT  
IYGLGLENGSKLQVTTFTDYNNLGSIIANNVTINDYVTLTGTGGTAGTDFDFAKITG  
SYNGNANRVFVDSFSPRSVATOKCTVYLGNALVNSIGSDTDPVAVRFTGN  
DSGAGLQGNISQNDIGTYNTLILNSVILGGGTAINGEIDLILNLIIFANGSTW  
GONTISLITLVSGNIGQVVIADQVNAATTTGTTTIKIQDNANANVSGTQAYTLIQ  
GAARNGTILGAPNAVTSIGTFAEYELIRDSNQDYVLTIRNDLVNTAVGNSAIAN  
APGVSONISRCLESTNTAAAYNNMLIAKDPDVAITVGTATDT' SAATVTVNNDTOKT





```
Query Match          9.1%; Score 18; DB 7; Length 1563;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  92 GAAAAATATATATTGG 109
    (((((((((((((((((((
Db  1522 GAAAAATATATATTGG 1539

RESULT 29
BNSLOGPB
LOCUS      BNSLOGPB      1563 bp      mRNA      PLN      29-JUL-1997
DEFINITION B.napus mRNA for endogenous S-locus glycoprotein (cone SLR1-WS-1).
ACCESSION  Z21609
VERSION     Z21609.1 GI:17817
KEYWORDS   endogenous S-locus glycoprotein gene.
SOURCE     rape.
ORGANISM   Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 1563)
AUTHORS   Oldknow,J., Franklin,T.M., Allard,S., Robert,L.S. and Trick,M.
TITLE     DNA sequences of the two homoeologous SLR1 genes of Brassica napus
JOURNAL   cv inestar
SEX       Plant, Reprod. 8, 254-255 (1995)
REFERENCE  2 (bases 1 to 1563)
AUTHORS   Robert,L.S.
TITLE     Direct Submission
JOURNAL   Submitted (05-FEB-1993) Robert L. S., Agriculture Canada, Plant
            Research Centre, 960 Carling Avenue, Ottawa, Ontario, Canada, K1A
            0C6
FEATURES   Location/Qualifiers
            source          1..1563
                        /organism="Brassica napus"
                        /strain="Westar"
                        /db_xref="taxon:3708"
                        /clone_lib="lambda gt 10"
                        /clone="SLR1-WS-1"
                        14..1354
                        /codon_start=1
                        /product="glycoprotein"
                        /protein_id="CAA79735.1"
                        /db_xref="GI:2285899"
                        /db_xref="SPTREMBL:Q39357"
                        /translation="MKGVIPVHHSYTFIFVILVLPFAFNTILSSNEALTISNKT
                        LVSQPVFELGFEKATKNSQSGSDRAVLGIWYKTTGDKTYVYVNRDNLPLNSIG
                        TLKISASLVLDQSDTTVWSNLGVAHLPVTAELLANGFVLKDSKINDLDRFMQ
                        SEDYPVDTLPEKIGRKHSEKEKILTSWKSPTDPSSGDYSLILETEGLHEFYLF
                        KNEFKYRTGPNVGRFNGIPKMQNWSYIDNSFIDNNEEYAYTFKVINNNMTIHSRF
                        RMSSTGYLOVITWTKTVPORNMFWSPEDCDLYKVCOPYAYCDMHTSPTCNCIKGFVPK
                        PNAGRWLDMSGGCVRSKLSGCGDFLRMSQKLPETSEALVEKRIGLKECREK
                        CYRDCNCTGYANMDIMDGGSGCVTWTGELVDNRKYDAGQDLYKVAEASLVPS"

BASE COUNT  483 a 306 c 349 g 425 t
ORIGIN

Query Match          9.1%; Score 18; DB 7; Length 1563;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  92 GAAAAATATATATTGG 109
    (((((((((((((((((((
Db  1522 GAAAAATATATATTGG 1539

RESULT 30
BNSLIR
LOCUS      BNSLIR      1794 bp      DNA      PLN      09-FEB-1996
DEFINITION B.napus (Westar) SLR1 gene for SLR1 glycoprotein.
ACCESSION  Z26914
```

```
VERSION  Z26914.1 GI:456316
KEYWORDS SLR1 gene; SLR1 glycoprotein.
SOURCE   rape.
ORGANISM Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 1794)
AUTHORS   Oldknow,J.
TITLE     Direct Submission
JOURNAL   Submitted (13-OCT-1993) Oldknow J., Cambridge Laboratory, Institute
            of Plant Science Research, Brassica & Oilseeds Research, John Innes
            Centre, Colney Lane, Norwich, Norfolk, UK, NR4 7UJ
REMARK    revised by [3] MAT
REFERENCE  2 (bases 1 to 1794)
AUTHORS   Oldknow,J.
TITLE     Direct Submission
JOURNAL   Submitted (22-FEB-1994) Oldknow J., Cambridge Laboratory, Institute
            of Plant Science Research, Brassica & Oilseeds Research, John Innes
            Centre, Colney Lane, Norwich, Norfolk, UK, NR4 7UJ
REFERENCE  3 (bases 1 to 1794)
AUTHORS   Oldknow,J., Franklin,T.M., Allard,S., Robert,L.S. and Trick,M.
TITLE     DNA sequences of the two homoeologous SLR1 genes of Brassica napus
JOURNAL   cv inestar
SEX       Plant, Reprod. 8, 254-255 (1995)
COMMENT   On Feb 23, 1994 this sequence version replaced gi:407774.
FEATURES   Location/Qualifiers
            source          1..1794
                        /organism="Brassica napus"
                        /strain="Westar"
                        /db_xref="taxon:3708"
                        /clone="2a2"
                        /tissue_type="Stigma"
                        /clone_lib="Lambda gtl0"
                        142..1476
                        /gene="SLR1"
                        142..1476
                        /gene="SLR1"
                        /citation={3}
                        /codon_start=1
                        /evidence=experimental
                        /product="SLR1 glycoprotein"
                        /protein_id="CAA81540.1"
                        /db_xref="GI:456317"
                        /db_xref="SPTREMBL:O39356"
                        /translation="MRGVIPVHHSYTFIFVILVLPFHVFTNTLSPNENLTISNKN
                        TLVSPGVFELGFEKTTTRNSPDGTRVAHLPVTAELLANGFVLKDSKINDLDRFIWO
                        TLKISASLVLDHSTNPWSTNFTGVAHLPVTAELLANGFVLKDSKINDLDRFIWO
                        SEDYPVDTLPEKIGRNLIGSENEKILTSWKSPTDPSSGDYSLILETEGLHEFYLL
                        KNEFKYRTGPNVGRFNGIPKMQNWSYIDNSFIDNNEEYAYSFQVNNNNHIHFRFM
                        SSTGYLOVITWTKTVPORNMFWSPEDCDLYKVCOPYAYCDMHTSPTCNCIKGFVPK
                        NAGRWLDMSGGCVRSKLSGCGDFLRMSQKLPETSEAVVDKRIGLKECREKV
                        RCNCTGYANMDIMDGGSGCVMTGELDDMRKYNAGGQDLYKVAASLVPS"
                        142..225
                        /gene="SLR1"
                        236..1473
                        /gene="SLR1"
                        /product="SLR1 glycoprotein"
                        1538..1543
BASE COUNT  566 a 351 c 381 g 496 t
ORIGIN

Query Match          9.1%; Score 18; DB 7; Length 1794;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  92 GAAAAATATATATTGG 109
    (((((((((((((((((((
Db  1642 GAAAAATATATATTGG 1659
    (((((((((((((((((((

RESULT 31
```

```

MMU13839      1825 bp      mRNA      ROD      03-DEC-1998
LOCUS      Mus musculus vacuolar adenosine triphosphatase subunit C mRNA,
DEFINITION      complete cds.
ACCESSION      U13839
VERSION      U13839.1 GI:3955097
KEYWORDS      .
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE      1 (bases 1 to 1825)
AUTHORS      Howell, M.L. and Dean, G.E.
JOURNAL      CDNA sequences for mouse vacuolar ATPase subunits
REFERENCE      2 (bases 1 to 1825)
AUTHORS      Dean, G.E.
TITLE      Direct Submission
JOURNAL      Submitted (19-AUG-1994) Gary E. Dean, Mol. Genetics, Biochemistry,
and Microbiology, University of Cincinnati College of Medicine, 231
Bethesda Avenue, Cincinnati, OH 45267-0524, USA
FEATURES      Location/Qualifiers
source      1..1825
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /cell_line="ES-D3"
             /dev_stage="18.5 day"
             156..1304
             /function="vacuolar acidification"
             /note="subunit C, V-ATPase"
             /codon_start=1
             /evidence=experimental
             /product="vacuolar adenosine triphosphatase subunit C"
             /protein_id="AAC83084.1"
             /db_xref="GI:3955098"
             /translations="MTEFWLISAPGKTCQOTWEKLAHTKNNLAVSKENIPDLK
             VGLDVLVGSDELAKLDAFEGVGVKQVAGYADVLEDSKDKVQENLLASGVDLVYII
             TRQWDWARKPIKSLNKISIEIIAGVTQIDDLKRSAGYNLKNLQNLQNKKNAGS
             LLRSLEAIIKDFDVLDSYLVLLVVPKLNHDWIKOYETLAEMVYPRSNVLSL
             DQSYLCNVTLFRKAVDDFRHAKENKFIVRDFQYNEEMKADKEMTLSDKKKQF
             GPLVRLKLVNFSFAFIAMIHKALRVFESVLRGLPVNFQAMLLQPNKSKVKLLREV
             LHELYKHLDSAAAIIDAPMDIPCLNLSQEQEYYPYVYIKDCNLLFEK"
BASE COUNT      521 a 371 c 434 g 499 t
ORIGIN
Query Match      9.1%; Score 18; DB 12; Length 1825;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 AGAAAAATATATATTG 108
      |||||
Db 1456 AGAAAAATATATATTG 1473

RESULT 32
BOSLRG      2139 bp      DNA      PLN      23-MAR-1995
LOCUS      Brassica gene for S locus related glycoprotein.
DEFINITION      X52089
ACCESSION      X52089.1 GI:117896
VERSION      S locus related glycoprotein: secreted protein.
KEYWORDS      Brassica oleracea.
SOURCE      Brassica oleracea
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Caprales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 2139)
AUTHORS      Trick, M.
TITLE      Direct Submission
JOURNAL      Submitted (14-MAR-1990) Trick M., Cambridge Laboratory, John Innes
Centre for Plant Science Research, Colney Lane, Norwich NR4 7UJ, U
K

```

```

REFERENCE      2 (bases 1 to 2139)
AUTHORS      Trick, M.
TITLE      Genomic sequence of a Brassica S locus-related gene
JOURNAL      Plant Mol. Biol. 15 (1), 203-205 (1990)
MEDLINE      91355861
COMMENT      Data kindly reviewed (02-JUN-1990) by Trick M.
FEATURES      Location/Qualifiers
source      1..2139
             /organism="Brassica oleracea"
             /strain="albobolabra"
             /db_xref="taxon:3712"
             /cell_line="plant line S63"
             /clone_lib="Lambda EMBL4"
             /clone="BS63-1"
             220..226
             /note="TATA box"
             329..335
             /note="TATA box related"
             379..1713
             /note="S locus related glycoprotein precursor (AA -28 to
             416)"
             /codon_start=1
             /protein_id="CAA36307.1"
             /db_xref="GI:17897"
             /translation="MRGVIPNYHHSYTLFFVILVLPFHVSTNTLSPNALTISSNK
             TLVSPGVDFELGFKTTNRNSPDGTDWILGYKTTSGHRTYVYANRDNALHNSMG
             TKISHASLVLLDHSNTPTVNSTFTGVAHLPTVAELLANGFVLURDSKTNLDLRFMMQ
             SFDPVDTLLPEMKLGRNLIGSENEKILTWSKSPDPSGDGDFSLTEGLHEFYLL
             KNEEKYVTPGWVRENGIPKMQMWSYIDNSFIDNNEVAYSFQVNNNNIHTFRFM
             SSTGLQVITWTKTVPQNMFWSPEDCDLTKVCGPIACDMDHSTPCNICIKGFVFK
             NAGWRDMDSGGCVRSKSLSCGEGDGLRMSOMKLPETSEAVVDRKIGLKECKRCV
             RDCNCTGYANNDDIMNGSGCVMTGELDDMRKYNAGGODLYVKVAAASLVPS"
             379..462
             /product="signal peptide (AA -28 to -1)"
             463..1710
             /product="mature peptide (AA 1 to 416)"
             1775..1780
             /note="polyA signal"
BASE COUNT      680 a 409 c 446 g 604 t
ORIGIN
Query Match      9.1%; Score 18; DB 7; Length 2139;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 92 GAAAAATATATATTG 109
      |||||
Db 1879 GAAAAATATATATTG 1896

RESULT 33
PFSTARP/C      2787 bp      DNA      INV      11-MAY-1995
LOCUS      P.falciiparum gene for STARP antigen.
DEFINITION      226314
ACCESSION      226314.1 GI:499324
VERSION      STARP antigen.
KEYWORDS      malaria parasite P. falciiparum.
SOURCE      Plasmodium falciiparum
ORGANISM      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE      1 (bases 1 to 2787)
AUTHORS      Fidock, D.A., Bottius, E., Brahimi, K., Moelans, I.M.D., Aikawa, M.,
Konings, R.N.H., Certa, U., Olafsson, P., Kaidoh, T., Asavanich, A.,
Guerin-Marchand, C. and Drulhe, P.
TITLE      Cloning and characterization of a novel Plasmodium falciiparum
sporozoite surface antigen, STARP
JOURNAL      Mol. Biochem. Parasitol. 64 (2), 219-232 (1994)
MEDLINE      95021499
REFERENCE      2 (bases 1 to 2787)
AUTHORS      Fidock, D.A.
TITLE      Direct Submission

```



```

exon      2288..2678
/feature="rpl15"
/citation={2}
/number=4
/feature="experimental"
2519..2678
/feature="rpl15"
/citation={2}
/feature="experimental"
2678..2680
/citation={2}
/feature="experimental"
BASE COUNT  920 a 522 c 550 g 992 t
ORIGIN
Query Match      9.1%; Score 18; DB 7; Length 2984;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  94 AAAATATATATATGGTG 111
|||||
Db  798 AAAATATATATATGGTG 781

RESULT 35
LOCUS      AF102623      7186 bp      DNA      BCT      24-MAY-1999
DEFINITION Methanosarcina barkeri dimethylamine corrinoid protein Mtbc (mtbc),
            trimethylamine methyltransferase Mtbt (mtbt), trimethylamine
            corrinoid protein Mtct (mtct), putative transmembrane protein Mtpt
            (mtpt), and dimethylamine methyltransferase Mtbb1 (mtbb1) genes,
            complete cds.
ACCESSION  AF102623
VERSION     AF102623.1 GI:4262423
KEYWORDS   Methanosarcina barkeri.
SOURCE     Methanosarcina barkeri.
            Archaea; Euryarchaeota; Methanosarcinales; Methanosarcinaceae;
            Methanosarcina.
REFERENCE  1 (bases 1 to 7186)
AUTHORS    Ferguson,D.J., Jr. and Krzycki,J.A.
TITLE      Reconstitution of trimethylamine-dependent coenzyme M methylation
            with the trimethylamine corrinoid protein and the isozymes of
            methyltransferase II from Methanosarcina barkeri
JOURNAL    J. Bacteriol. 179 (3), 846-852 (1997)
MEDLINE    97158682
REFERENCE  2 (bases 1 to 7186)
AUTHORS    Ferguson,D.J., Gorlatova,N., Paul,L., Grahame,D. and Krzycki,J.A.
TITLE      The corrinoid protein from Methanosarcina barkeri specific for
            dimethylamine: CoM methyl transfer
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 7186)
AUTHORS    Paul,L. and Krzycki,J.A.
TITLE      The genes encoding the trimethylamine and dimethylamine
            methyltransferases of Methanosarcina barkeri are cotranscribed and
            interrupted by translationally bypassed in-frame amber codons
JOURNAL    Unpublished
REFERENCE  4 (bases 1 to 7186)
AUTHORS    Paul,L. and Krzycki,J.A.
TITLE      Direct submission
JOURNAL    Submitted (27-Oct-1998) Microbiology, Ohio State University, 484
            West 12th Ave, Columbus, OH 43210, USA
FEATURES   Location/Qualifiers
            1..7186
                /organism="Methanosarcina barkeri"
                /strain="MS"
                /db_xref="taxon:2208"
                734..1375
                /gene="mtbc"
                734..1375
                /gene="mtbc"
                /function="participates in dimethylamine:coenzyme M methyl

```

```

transfer"
/feature="ORF identified by N-terminal sequence of isolated
corrinoid protein"
/codon_start=1
/transl_table=11
/product="dimethylamine corrinoid protein Mtbc"
/protein_id="AA014629.1"
/db_xref="GI:4262424"
/translation="MSKEELIQELADAIISCKKDTVLAVVEKAKGELEPSEIEIKGLA
AGMNEVGVRFGKFLPHVMAADAMTAGAALKDLMPEGASGSLGVVNGTVEGD
VHDIGKATVSTMLQSGFEVHDIGRDVPIRNFIEKAEVNADMGISALMTTLQSQK
SVIELKEGLRDKRVKVMGAPATQAWADKIGADCYAENATEAVAKAKELLA"
1402..2889
/feature="mttb"
1402..2889
/feature="mttb"
/feature="DAG codon is translated by an undetermined
mechanism: ORF identified with N-terminal sequence of 52
kDa trimethylamine methyltransferase polypeptide"
/codon_start=1
/transl_table=11
/feature="except-(pos:2401..2403,aa:OTHER)"
/evidence="experimental"
/product="trimethylamine methyltransferase Mtbt"
/protein_id="AA014630.1"
/db_xref="GI:4262425"
/translation="MAKNNAVAGFNALGVNELNLTDELKATHYATMEVLMDPGLOV
SDPARQIFRENGCEVNEKTVYKIPYLVKALQALPSFVILNGRDKKFNIVQBCGG
KVHTCFGTGKVKYQODGRYIVDSVERKDIAKDCDAENIDIFSLPVSARDIAG
QADQVHETLPLANTAKHFHIDPVENVEYRDIVKAYYGGDEEFAKRPFSMLL
CPTPLSLVNAQVVIKARFIPVNLVSNMSSGSPVYLAGTLVTHNAEVLGIV
LAQTVFGAKWYGSSTTDFDLKGTAPVSGPELGLISAALAKAQFVGLNPSYVAGS
SDAKVPDQAGHEKMTTLLPALAGANTYAGAGLELGMTFSMEQVLDIONDIFSVMKK
AMOGIPVSEETLAVESIQKVGIGNFLAKQTRQVLDYPSNMLLDRHMGDWAAGS
KDLATVAHEKVEDVLKHHQVTPIDAFKDMQAIVDKADKAFRM"
2920..3573
/feature="mttc"
2920..3573
/feature="mttc"
/function="participates in trimethylamine:coenzyme M
methyl transfer with TMA methyltransferase"
/feature="ORF identified by N-terminal sequence of corrinoid
protein"
/codon_start=1
/transl_table=11
/product="trimethylamine corrinoid protein Mtct"
/protein_id="AA014631.1"
/db_xref="GI:4262426"
/translation="MANKEELIAKAKEAITDFDELAEEVANEALAGIDPVELIEKG
FTAGMEVGEKFGOGELFLPHVLAEEAMNSGKIVTPMEKRSKTSKSLGTVAIGTI
EGDTHSGKDIIVASMLNAGEVYVDLGRDVPINTFEVKELKPOVAVASSALMTMV
NQDIEQLKAEAGVRDQVKTWGGAPVQDMADKIGADIYGESANDAVAKVKAALNV"
3885..4946
/feature="mttb"
3885..4946
/feature="mttb"
/function="possible methylamine permease"
/codon_start=1
/transl_table=11
/product="putative transmembrane protein Mtpt"
/protein_id="AA014632.1"
/db_xref="GI:4262427"
/translation="MEGVFVLDLAKAKQESKRISKGYMAFFCAIFWIGYLPVGTIV
WLNPFDMSSAIKTSKDGISLVVTAITAFNALVLMVWNGVIGKYGELVRT
LKEFPCSKVFFLASIFGGMATLIGSFIAFGGFAAVAGLLYPVGSLIAYKYG
EKISRAAIGIAVILGSIYGGGLFTELSSGNVPKIGYLAGLMAAGWEGTAG
KGLDIAEDAGLILRLGCEINIIWIIIVPLALVGPMSYFALQDFDPTLILLFAG
ITFGCIYTWKISFLIGVGRGOGIGNLYGCAIIFLFFGVDVPTWIIIGALUCIA
GSFINTEDASELETLRGE"
5470..6873
/feature="mtbb1"
5470..6873
/feature="mtbb1"

```



```
homolog="oligopeptide transport ATP-binding protein
/protein_id="CAA68076.1"
/db_xref="GI:2911536"
/translation="MAKIINKHDNDKNSDITIKHQKINKSTRVLDSIRVEPLKNNOI
LKRVNHVFPQGHKSTHVRGVDLDRHGEIIGLVGSGGSKSVTSKYLINNVESG
IISADQIQTVEELNIIKKQKYMVGIGKIGYPODPLTSLNPNRVGKOLLALNN
NEDKWNSTTEKKNYILGLKTFGRLSAREKIFAYPHILSGCMQKQRIIVIMVALPE
LIADSPITTAOPTVOASVALFEDIRQMGISILISHSNISVIAKECDIYVMIAGR
IVEKQKQIFNPHAPHTWALISAIPESKDEKLYTQGTDPDMANFLPGDPFYRND
YALEIDYLKEPPLIPITTHAAATWLLSPEAPKIEKSELOKRLDLFEKKAFFNE"
7114..9642
/gene="oppF"
CDS
7114..9642
/gene="oppF"
/codon_start=1
/transl_table=4
/product="Oligopeptide transport ATP-binding protein
homolog"
/protein_id="CAA68077.1"
/db_xref="GI:2911537"
/translation="MNNKFKDKKVLISDNLKKYFVNOGMINKAVDGVSEDEVHEEI
VLIGSGSGKTIVTSTILRLYDDYNGVFKRLDIISGRISSELNFLRKNVOMIFQ
DPHASUNQOITSILKEPLLVNGYMKDKIKDIFDNLNKKNFKYTFQIEARMLEIQ
NYLEINKLAEPFLKWTNKNLEFNDFISREDNFSFFGYLEEKQNVESIINMYVA
NVDRIEFYETORRYRNKELPGQOIALNEKEELKYTOCLSKTSKAYEASILEKVEY
KQIENLRHKIEKISSEKNTFANYIHESKEKSLIDIALRLMSCDLDFSYNLSKELL
MKRGVOIAKIEKSEKNTGFNNKILINELDEYINEFYKTLAPLYSKYLSIDIKNY
LEKDPFKDEYTKISELEKQLDSELQKLSREVEIKVINKENPNEITSEQIQEAK
NDELGQKYLSEGRKILISYKAELESYLNLEBOKKPYLELRKQDDYCNKKYKELQK
KFEYVEQLRQEKILYKTAELQKQNDKQKSLAKLSKLSIKYQNYITLKM
YHSDLSLKEDLKFEDIERKYLDKINNYVLGTDHKWETNLTADAMAGVDAXHM
KWKRRYDVSXSVPIAKLLISLYKTIIYKALEDVGLLKQFAYRPHFSGGOLQ
RIVARALIVEQVIADEPIASLDISIOQAVNLLKELCIKKKIGLIFIAHDLISME
YVADEVQIMHLCKIVESGKTEIAYNPIHYTINLFKAIPKISNANEKFNQVSEALDY
LDEQFPNPPEFKGEBHYVYGTSOLKQKTHNAKLEKEHEL"
BASE COUNT 4157 a 1415 c 1501 g 3341 t
ORIGIN

Query Match 9.1% Score 18; DB 1; Length 10414;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 AACATTAGGGAATTTTG 70
|||||
Db 9534 AACATTAGGGAATTTTG 9517

RESULT 37
CEM18
LOCUS CEM18 26542 bp DNA INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid M18, complete sequence.
ACCESSION Z68507
VERSION 268507.1 GI:159944
KEYWORDS HTG: Collagen; DNA repair protein like; Dyenin like.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favella,A.,
Fulton,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
Saunders,D., Showkhen,R., Smalton,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
```

```
TITLE
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 26542)
AUTHORS Steward,C.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1996) Louis, MQ 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webace.sanger.ac.uk/cgi-
bin/display?db=wormace&class=Sequence&object=M18
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequence
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: this sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone M18. It
may be shorter because we only sequence overlapping sections once,
or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone M18 is at 1 in this sequence. The true
right end of clone M18 is at 10009 in
sequence Z69634.
The true left end of clone B0001 is at 26439 in this sequence. The
true right end of clone F11A10 is at 23200 in this sequence. The
start of this sequence (1..104) overlaps with the end of sequence
Z68297.
The end of this sequence (26439..26542) overlaps with the start of
sequence Z69634.
FEATURES
Location/Qualifiers
1..26542
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="M18"
5816..8308
/gene="M18.3"
join(5816..6232,7147..7315,7361..7494,7545..7993,
8050..8308)
/gene="M18.3"
/note="predicted using GeneFinder: cDNA EST EMBL:D73035
comes from this gene; cDNA EST EMBL:D76008 comes from this
gene; cDNA EST YK587dl.3 comes from this gene"
/protein_id="CAA932825.1"
/db_xref="GI:3878705"
/codon_start=1
/db_xref="SPTREMBL:021555"
/translation="MGIDYNRMWHLKDTLGEELKGVKIMBEDLYVSGTTSQHLVHS
KLNSGYSNIVNEILEQQGWSIIYNNQKLKLEIGRDVFGROSLVFNESMIFGCRIT
APTSKGWIEPFGQVTVFLSDSDPVYISILPEYENVLGDFYSGSDISRIYVNRKH
GLNVSRSTGGEIDSFNTKDLQKIDATKILKNYHSHVAVCLSGGVDSTVAHVH
HSSYPENQIDILINVAFGSEKDEQAPDRKARALESEFTAYPTRFQRLILNVNDS
OTLEYNRKESIDAAQPASSVLDLSCSLVFAVRAEGVDSNQSVRSPATCTLLGS
GADELLAGYARHRTFRFEQIIPENVAEECNELRLGSRNGRDARVAALQGTILSP
LLEDTVTWLNALPVDKSDWLSLPRVGEGKQLLRITVKMLGSPYDAPKAMQFGRMA
KMSNAGDSIKSGSKSPLHLKTDVN"
complement(8729..10202)
/gene="M18.1"
complement(join(8729..9598,10059..10202))
/gene="M18.1"
/note="predicted using GeneFinder: similar to collagen;
cDNA EST EMBL:D70719 comes from this gene; cDNA EST
EMBL:D66790 comes from this gene; cDNA EST EMBL:D65910
```



TITLE The sequence of C. elegans cosmid F46H5  
 JOURNAL Unpublished (1995)  
 REFERENCE 3 (bases 1 to 3886)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-NOV-1995) Robert Waterston  
 REFERENCE 4 (bases 1 to 3886)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-NOV-1996)  
 COMMENT On Nov 29, 1996 this sequence version replaced gi:1109877.  
 Submitted by:  
 Genome Sequencing Center  
 Department of Genetics, Washington University,  
 St. Louis, MO 63110, USA, and  
 Sanger Centre, Hinxton Hall  
 Cambridge CB10 1RQ, England  
 e-mail: rwnematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.  
 It may be shorter because we only sequence overlapping sections  
 once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate  
 chemistry; an attempt was made to resolve all sequencing problems,  
 such as compressions and repeats; all regions were covered by  
 sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 5' cosmid is ZK967, 200 bp overlap; 3' cosmid is T10A3, 800 bp  
 overlap. Actual start of this cosmid is at base position 1 of  
 CELF46H5; actual end is at 38286 of CELF46H5

#### NOTES:

Coding sequences below are predicted from computer analysis, using  
 the program Genefinder (P. Green and L. Hillier, ms in preparation).  
 Location/Qualifiers  
 1. 38886

FEATURES  
 source  
 /organism="Caenorhabditis elegans"  
 /strain="Bristol N2"  
 /db\_xref="taxon:6239"  
 /chromosome="X"  
 /clone="F46H5"  
 complement(225. .1802)  
 /gene="F46H5.3"  
 complement(join(225. .714,767. .1092,1430. .1636,1692. .1802))  
 /gene="F46H5.3"  
 /note="Similar to other ATP:guanoide phosphotransferases;  
 coded for by C. elegans cDNA yk61c6.3; coded for by C.  
 elegans cDNA yk119h1.3; coded for by C. elegans cDNA  
 yk68a3.3; coded for by C. elegans cDNA CEMS843F; coded for  
 by C. elegans cDNA yk116b11.5; coded for by C. elegans  
 cDNA yk61c6.5; coded for by C. elegans cDNA cml14a3; coded  
 for by C. elegans cDNA CEMS1414; coded for by C. elegans  
 cDNA yk68a3.5; coded for by C. elegans cDNA yk82d10.5;  
 coded for by C. elegans cDNA CEMS86R; coded for by C.  
 elegans cDNA yk124f10.5; coded for by C. elegans cDNA  
 yk119h1.5; coded for by C. elegans cDNA CEMS86F; coded  
 for by C. elegans cDNA CEMS86FB; coded for by C. elegans  
 cDNA yk142e5.3; coded for by C. elegans cDNA yk142e5.5;  
 coded for by C. elegans cDNA yk162g7.3; coded for by C.  
 elegans cDNA yk162g7.5; coded for by C. elegans cDNA  
 yk174e8.3; coded for by C. elegans cDNA yk174e8.5; coded  
 for by C. elegans cDNA yk184d9.3; coded for by C. elegans  
 cDNA yk184d9.5"  
 /codon\_start=1

gene  
 CDS  
 /protein\_id="BAB37022.1"  
 /db\_xref="GI:1109878"  
 /translation="MLREAYSTSLHCVQAEQMSVSPDVIAKIEBYAKLQAAPECHS  
 LKLYLTREVVDLQDKRTKLGALNDVIQSVGNSLVGAVYAPDAEATYTLKPLFD  
 PLIOYHNGFAPDAKQPNIDEGKTALYDLDPKGFINSTIRSGRSLQGYFNPNC  
 LSEANYLEMESKVAIFDNTIDPELAGKYFPLDGMTKEIQDLIKDHFLEKGGDFLQ  
 AANACRYWPKRGIFNPKTFLWCNEEDHLRIISMOEGNGVGVLERLIKGVKTE  
 KOAPESRDDRLGWLTFPCNSLGTIVRASVHIRLPKISAKPDFKSIDCGLKLQIRGHH  
 EHSSEGGVYDISNKAIRLGLTEFEAVKQMDGIAHLIALEKAA"  
 complement(8345. .12126)  
 /gene="F46H5.7"  
 complement(join(8345. .8450,8498. .8586,8630. .8789,  
 8840. .8920,9422. .9643,9693. .9914,9964. .10083,10136. .10297,  
 10356. .10572,11139. .11231,11324. .11397,11884. .12022,  
 12093. .12126))  
 /gene="F46H5.7"  
 /note="Coded for by C. elegans cDNA yk100c6.3; coded for  
 by C. elegans cDNA yk100c6.5; coded for by C. elegans cDNA  
 yk134h9.5; coded for by C. elegans cDNA cm08f2; coded for  
 by C. elegans cDNA cm20g12"  
 /codon\_start=1  
 /protein\_id="BAB37024.1"  
 /db\_xref="GI:1109879"  
 /translation="MADVATKIPDANVRLEIEPVQVSHDQDATLNFATVQSVPGAH  
 GUYRDGGERKALLFDNNTGKVPYPPNGWDSVPFIHLAHSRHHGTHFYADYKANDFA  
 ENISAVOKFLAAGLGRGSKVVAACKTRARSTKKNQPSRQFDAQPENVAISESE  
 GRDSEKENDRQLEKLRNDELKIRVKEYKSLQSAQLKVKLEKLSAQASDSE  
 KFYRDSIDKRSQDHDEINNHVIEELRNKLATIVESKNKDLTKTSESNQDILKN  
 AKDYTFLENLQNSDAHEEVKSTAVRALEVLGLANSTIRQAEAKQQLQEANWYANE  
 RVGKLEQNGYLGKITQLKARADTSHAEXMLKDKSEKRVWEINEEKSKLEWRLQSQ  
 WMDAKRWGELSSVALORNLDTANSKIQSLNDQSHSSTMTPTDGTIISOSNOGT  
 WNLANGNPQCFGTGAIPIPLVSALPFGANPFIIFGANGRNVTIISNTFOEVLVT  
 GSFINWKTCKEKLVSCKKGVTVLTRGRHEFRINGEWATSSDYQQVQVNLGSGN  
 NIIFVE"  
 complement(14754. .23897)  
 /gene="F46H5.4"  
 complement(join(14754. .15029,15678. .15873,15926. .16021,  
 16070. .16299,16348. .16711,16766. .16981,17030. .17235,  
 17286. .17447,17491. .17673,17891. .18148,18201. .18359,  
 18406. .18703,18752. .18871,18925. .19098,19147. .19418,  
 19465. .19610,19655. .19735,19899. .20133,20184. .20474,  
 20522. .20660,20711. .20820,20865. .21061,21113. .21316,  
 21360. .21521,21567. .21716,21764. .21884,21940. .22000,  
 22060. .22214,22261. .22448,22513. .22706,22827. .23352,  
 23400. .23486,23535. .23638,23684. .23792,23851. .23897))  
 /gene="F46H5.4"  
 /note="Similar to Rat trig gene product; coded for by C.  
 elegans cDNA yk31e7.5; coded for by C. elegans cDNA  
 yk40d6.5; coded for by C. elegans cDNA yk31e7.3; coded for  
 by C. elegans cDNA yk40d6.3; coded for by C. elegans cDNA  
 yk149g5.3; coded for by C. elegans cDNA yk149g5.5; coded  
 for by C. elegans cDNA yk153h2.3; coded for by C. elegans  
 cDNA yk153h2.5"  
 /codon\_start=1  
 /protein\_id="BAB37023.1"  
 /db\_xref="GI:1109880"  
 /translation="MKPADGARFVKRKSVTASEVRKHKVISCGLHPIHRLSGDNLGD  
 LAVSMERKIQTEPPVPVVDVFDLQDKRTCLLYSISPKSKKFLFYSDDEVETIR  
 ODGLDHYTTIPNIENHVRDICGYDNFSLVNRKYAQLGTEIDKRWNLKLTALR  
 SLRPIQFHNGIRIDREPSITIDGSDFTDKATEQFVYVSMKKSADIKNTLLY  
 SMLSKNRYKFMKNCINEESKWKRGVRLPEQETPKLFVKYKAAADPFFFLP  
 ASMAIYDKNRQVTESMFYNIADHLDMLGSHOPNFNNYMQVLFNTVGLKEDNFL  
 VYIKEVQLQNDVFENSEPTGTGDKENMERLEPRAEKNQRLGAVRSLGQVIDLQ  
 RYKANVSTGASSDRITDPMMSQCTASGAVITTAGOSQDCSITADASTIASM  
 GSTRKSGGSFADRIYFVRVTPITKRFKAVSNLPTQSEVPIENFSCNLKFSF  
 IQEGDKTSDSDIRICSEMRNTNGKVKHKNFPELITLAGNSKSKYQSHGSDNLT  
 NSERVYHAMEIPYQASLNKYNKVFYYPKHINSRNTGNARNITKIELMDANET  
 AOEVFENGSRMGLFTSAKTSYTHNRTPHFTDEIKLPLCDLNDHLLFTVYHS  
 CKEGSSSTESPIGTWMLPYRNKGLRSGNHLFVCGEKPYPRIYLDANNALPNLKW  
 VNHKPIFSCSTEVISVHAODEFLENFAGVASLNSNDPRKPPVGETOLIRSLGLEC  
 KTEPKKLTAFTIEMISRLLLFTIANPPYSDLSMKAFEYIGCLKLFENVLHLDLDAQ  
 RNMLVSVYKROKAAQESKPHSNTRPVELKSSPTDNLSSMIENHVERTHSAVNTGT  
 KNIRLHCLLEWLNRARGLSDVSLVHSWFLLEILLKSCSEYLTMTGRITHSPKRSFE



EQFLKNLETLVDILAQEVIRHTNDPKARMI SNSLGVFLRDCFSIMDRITFMKLVHK  
YLIAFAESMRKLVHNSLLSIKIDFVVVCSYEHYLI NVILSDLPKPSNTNGAVPPA  
SLFSGNRTKSSSTWTLNDISRTHYLSGOVLSOMKOSIIISGKTTLCAXAIEVVKEL  
LOSHDLDRIVEGESLAOVANIYKPLVTI VLNQIECLHSGSVRNSTDYSSNSTSVVEQT  
ORDVAAJAGKLRNSPDPEFGQONDLPMTKTIILCCVFWLVNDRDLKHWIRSLD  
NENKMLHILFHNTISPELKDDPASARSPKISLTLDPEPEGQVKYRAQSEUC  
ETKYQSTILEALSSAIVSECVFMCVIEVDNIATADPKNAOFHILPFIIMH  
GLSNASQDLEIVFAAQONFLAKFPDILEONPELCAELSQOILRHCSSTRLENVRT  
MAVSYLHFLRENFYLRNLTRATFLSTALSTLSSGCGIDIFYNDEFMIRSLIEIAN  
QLAADDFTDIAAKKLTQMOELTANLQIMLSTVRMREHVNDYEMTIDLMYOLVEG  
YSNPDLRI TWLMAERHEKORNLCEAAHSYQSAIVFEYIAQDONLSPESKGA  
TFSEITPNALIKESKTNFNSLKNADSENHICYSHFTEAGVIKILERAFALLEKAOLYEL  
LFPSKILKYVHATKSYSPVSHTRKLGIAADOIKETGEYENQSDAWISPLGIDK  
RCFGSFRVAFYKLGALNNAEYFKESAFKLEINSRLNETFTYNNIGGNVYVLK  
DSVPQLEKUNPEKAYIQITFDVYLSNEMKERTYTFRRNNVRYFYEPAPYMEGR  
AQGLAAQYKRRILTVENSFPYKTRQLQVNRSVKDFSPIEVAJEDIEKKTRELSAA  
AQHNKPMLESMFQGGSIGTVNGPLEIANVELANMLDGRPRVDRILQNLKRLSFRH  
LOCKMAEIELCRLQIGEDQKEYQORNVENEFSEFVTHLKPMLSRQNETTISEFSDPT  
VV"

gene complement(25261..26312)

/gene="F46H5.5"  
CDS complement(join(25261..25392,25442..25581,25632..25722,  
25768..26131,26185..26312))  
/gene="F46H5.5"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAB37026.1"

Query Match 9.1%; Score 18; DB 34; Length 38886;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 CACTAGAAAATATATAT 104  
|||||  
Db 7478 CACTAGAAAATATATAT 7461

RESULT 39  
AC008437/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CIT-HSPC\_329A22, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 31 unordered pieces.  
ACCESSION AC008437  
VERSION AC008437.1 GI:5686629  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 41402)  
DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 41402)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
www.jgi.doe.gov.  
COMMENT  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 31 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 843: contig of 843 bp in length  
gap of unknown length  
\* 844 1684: contig of 841 bp in length  
gap of unknown length  
\* 1685 2545: contig of 861 bp in length  
gap of unknown length  
\*

3353: contig of 808 bp in length  
gap of unknown length  
4230: contig of 877 bp in length  
gap of unknown length  
5091: contig of 861 bp in length  
gap of unknown length  
5092 6040: contig of 949 bp in length  
gap of unknown length  
6041 7557: contig of 1517 bp in length  
gap of unknown length  
7558 8758: contig of 1201 bp in length  
gap of unknown length  
8759 9648: contig of 890 bp in length  
gap of unknown length  
9649 10732: contig of 1084 bp in length  
gap of unknown length  
10733 11784: contig of 1052 bp in length  
gap of unknown length  
11785 12650: contig of 866 bp in length  
gap of unknown length  
12651 13529: contig of 879 bp in length  
gap of unknown length  
13530 14784: contig of 1255 bp in length  
gap of unknown length  
14785 16385: contig of 1601 bp in length  
gap of unknown length  
16386 17197: contig of 812 bp in length  
gap of unknown length  
17198 18040: contig of 843 bp in length  
gap of unknown length  
18041 19499: contig of 1459 bp in length  
gap of unknown length  
19500 20458: contig of 959 bp in length  
gap of unknown length  
20459 21887: contig of 1429 bp in length  
gap of unknown length  
21888 23364: contig of 1477 bp in length  
gap of unknown length  
23365 25150: contig of 1786 bp in length  
gap of unknown length  
25151 27026: contig of 1876 bp in length  
gap of unknown length  
27027 28817: contig of 1791 bp in length  
gap of unknown length  
28818 30556: contig of 1739 bp in length  
gap of unknown length  
30557 32533: contig of 1977 bp in length  
gap of unknown length  
32534 34180: contig of 1647 bp in length  
gap of unknown length  
34181 36262: contig of 2082 bp in length  
gap of unknown length  
36263 38508: contig of 2246 bp in length  
gap of unknown length  
38509 41402: contig of 2894 bp in length.  
FEATURES  
Location/Qualifiers  
source 1..41402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CIT-HSPC\_329A22"  
BASE COUNT 12577 a 7525 c 7689 g 13482 t 129 others  
ORIGIN  
Query Match 9.1%; Score 18; DB 41; Length 41402;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 118 ATAATTTTGAGATAATTA 135  
|||||  
Db 16210 ATAATTTTGAGATAATTA 16193

```

RESULT 40
AC016648/c
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone RPCI-1_137K24, HTG 04-DEC-1999
PROGRESS ***, 40 unordered pieces.
ACCESSION
AC016648
VERSION
AC016648.1 GI:5524058
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 45617)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1042: contig of 1042 bp in length
gap of unknown length
1043 1777: contig of 735 bp in length
gap of unknown length
1778 2698: contig of 921 bp in length
gap of unknown length
2699 3417: contig of 719 bp in length
gap of unknown length
3418 4250: contig of 833 bp in length
gap of unknown length
4251 4988: contig of 738 bp in length
gap of unknown length
4989 6185: contig of 1197 bp in length
gap of unknown length
6186 7030: contig of 845 bp in length
gap of unknown length
7031 7964: contig of 934 bp in length
gap of unknown length
7965 8740: contig of 776 bp in length
gap of unknown length
8741 9732: contig of 992 bp in length
gap of unknown length
9733 10862: contig of 1130 bp in length
gap of unknown length
10863 11583: contig of 721 bp in length
gap of unknown length
11584 12502: contig of 919 bp in length
gap of unknown length
12503 13207: contig of 705 bp in length
gap of unknown length
13208 14250: contig of 1043 bp in length
gap of unknown length
14251 15560: contig of 1310 bp in length
gap of unknown length
15561 16481: contig of 921 bp in length
gap of unknown length
16482 17643: contig of 1162 bp in length
gap of unknown length
17644 18527: contig of 884 bp in length
gap of unknown length
18528 19624: contig of 1097 bp in length
gap of unknown length
19625 20712: contig of 1088 bp in length

```

```

* 20713 21946: gap of unknown length
* contig of 1234 bp in length
* 21947 23095: gap of unknown length
* contig of 1149 bp in length
* 23096 24350: gap of unknown length
* contig of 1255 bp in length
* 24351 25672: gap of unknown length
* contig of 1322 bp in length
* 25673 26629: gap of unknown length
* contig of 957 bp in length
* 26630 27703: gap of unknown length
* contig of 1074 bp in length
* 27704 28979: gap of unknown length
* contig of 1276 bp in length
* 28980 30321: gap of unknown length
* contig of 1342 bp in length
* 30322 31485: gap of unknown length
* contig of 1164 bp in length
* 31486 32601: gap of unknown length
* contig of 1116 bp in length
* 32602 34106: gap of unknown length
* contig of 1505 bp in length
* 34107 35775: gap of unknown length
* contig of 1669 bp in length
* 35776 37641: gap of unknown length
* contig of 1866 bp in length
* 37642 39269: gap of unknown length
* contig of 1628 bp in length
* 39270 40997: gap of unknown length
* contig of 1728 bp in length
* 40998 42877: gap of unknown length
* contig of 1880 bp in length
* 42878 43958: gap of unknown length
* contig of 1081 bp in length
* 43959 45617: gap of unknown length
* contig of 1659 bp in length.
FEATURES
source
1..45617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPCI-1_137K24"
BASE COUNT 13659 a 8800 c 8954 g 14163 t 41 others
ORIGIN
Query Match 9.1%; Score 18; DB 44; Length 45617;
Best Local Similarity 100.0%; Fred. No. 43;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 178 AAGAACTCTATTATATGA 195
|||||
Db 28111 AAGAACTCTATTATATGA 28094
|||||
RESULT 41
AC012518/c
LOCUS
DEFINITION
Homo sapiens clone RP11-512G13, HTG 30-OCT-1999
unordered pieces.
ACCESSION
AC012518
VERSION
AC012518.1 GI:6139063
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 57784)
AUTHORS
Muzny, D.M., Adams, C., Bailey, M., Barbaria, J., Blankenburg, K.,
Bodoc, B., Bock, J., Bowie, S., Brooks, A., Buhray, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, X., Donah-Rashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,

```

Forcum-Tansey, J., Frantz, P., Ganesh, R., Correll, J.H., Correll, L.L., Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hognes, M., Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Osval, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sucgang, R., Tabor, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 57784)  
Worley, K.C.

Direct Submission  
Submitted (29-OCT-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

NOTE: This is a 'working draft' sequence. It currently consists of 48 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1  
2443: contig of 2443 bp in length  
2444  
2494  
4658: contig of 2165 bp in length  
4659  
4709  
6857: contig of 2149 bp in length  
6858  
6907: gap of unknown length  
6908  
8900: contig of 1993 bp in length  
8901  
8950: gap of unknown length  
8951  
10882  
10931: gap of unknown length  
10932  
12724: contig of 1793 bp in length  
12725  
12774: gap of unknown length  
14539: contig of 1765 bp in length  
14588: gap of unknown length  
14590  
16298: contig of 1709 bp in length  
16299  
16348: gap of unknown length  
16349  
17854: contig of 1506 bp in length  
17855  
17904: gap of unknown length  
17855  
19409: contig of 1505 bp in length  
19410  
19459: gap of unknown length  
20940: contig of 1445 bp in length  
20954: gap of unknown length  
20955  
22380: contig of 1426 bp in length  
22381  
22430: gap of unknown length  
22431  
23849: contig of 1419 bp in length  
23850  
23899: gap of unknown length  
23900  
25287: contig of 1388 bp in length  
25337: gap of unknown length  
25338  
26711: contig of 1374 bp in length  
26712  
26761: gap of unknown length  
26762  
28117: contig of 1355 bp in length  
28118  
28167: gap of unknown length  
28168  
29457: contig of 1290 bp in length  
29507: gap of unknown length  
29508  
29508  
30793: contig of 1286 bp in length  
30794  
30843: gap of unknown length  
32113: contig of 1270 bp in length  
32114  
32163: gap of unknown length  
32164  
33419: contig of 1256 bp in length  
33420  
33469: gap of unknown length  
33470  
34675: contig of 1206 bp in length  
34676  
34725: gap of unknown length  
35930: contig of 1205 bp in length  
35931  
35980: gap of unknown length

35981  
37186  
37236  
38362: contig of 1127 bp in length  
38363  
38412: gap of unknown length  
38413  
39535: contig of 1123 bp in length  
39585: gap of unknown length  
39586  
40659: contig of 1074 bp in length  
40660  
40709: gap of unknown length  
41623: contig of 914 bp in length  
41673: gap of unknown length  
42528: contig of 855 bp in length  
42578: gap of unknown length  
43427: contig of 849 bp in length  
43477: gap of unknown length  
43478  
43478  
44322: contig of 845 bp in length  
44371: gap of unknown length  
44372  
45187: contig of 816 bp in length  
45188  
45236: gap of unknown length  
46042: contig of 806 bp in length  
46043  
46091: gap of unknown length  
46092  
46892: contig of 801 bp in length  
46893  
46941: gap of unknown length  
46942  
47740: contig of 799 bp in length  
47741  
47789: gap of unknown length  
48120: contig of 331 bp in length  
48121  
48169: gap of unknown length  
48170  
48504: contig of 335 bp in length  
48505  
48553: gap of unknown length  
48554  
49348: contig of 795 bp in length  
49349  
49397: gap of unknown length  
50192: contig of 795 bp in length  
50193  
50241: gap of unknown length  
50242  
51036: contig of 795 bp in length  
51037  
51085: gap of unknown length  
51086  
51877: contig of 792 bp in length  
51878  
51926: gap of unknown length  
51927  
52177: contig of 791 bp in length  
52178  
52766: gap of unknown length  
52767  
53557: contig of 791 bp in length  
53558  
53606: gap of unknown length  
53607  
54395: contig of 789 bp in length  
54396  
54444: gap of unknown length  
54396  
55232: contig of 788 bp in length  
54445  
55281: gap of unknown length  
55233  
55598: contig of 317 bp in length  
55599  
55647: gap of unknown length  
55648  
56426: contig of 779 bp in length  
56427  
56475: gap of unknown length  
56476  
57175: contig of 700 bp in length  
57176  
57224: gap of unknown length  
57225  
57784: contig of 560 bp in length.

FEATURES  
source

1..57784  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rp11-512G13"  
BASE COUNT 16242 a 11269 c 11244 g 16579 t 2450 others  
ORIGIN

Query Match g.1% Score 18; DB 42; Length 57784;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 AAAATATATATATGGTG 111  
|||||  
Db 32194 AAAATATATATATGGTG 32177

RESULT 42  
AC010280/c

AC010280 70390 bp DNA HTG 15-SEP-1999  
LOCUS Homo sapiens chromosome 5 clone CIT-HSPC\_537E7, \*\*\* SEQUENCING IN

PROGRESS \*\*\*, 84 unordered pieces.  
AC010280  
VERSION AC010280.1 GI:5882653  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 70390)  
REFERENCE DOE Joint Genome Institute.  
AUTHORS Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 70390)  
AUTHORS DOE Joint Genome Institute.  
JOURNAL Direct Submission  
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
www.jgi.doe.gov.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 84 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 596: contig of 596 bp in length  
gap of unknown length  
597 1213: contig of 617 bp in length  
gap of unknown length  
1214 1865: contig of 852 bp in length  
gap of unknown length  
1866 2428: contig of 563 bp in length  
gap of unknown length  
2429 3529: contig of 1101 bp in length  
gap of unknown length  
3530 4154: contig of 625 bp in length  
gap of unknown length  
4155 4933: contig of 779 bp in length  
gap of unknown length  
4934 5574: contig of 641 bp in length  
gap of unknown length  
5575 6224: contig of 650 bp in length  
gap of unknown length  
6225 5962: contig of 738 bp in length  
gap of unknown length  
6963 7608: contig of 646 bp in length  
gap of unknown length  
7609 8291: contig of 683 bp in length  
gap of unknown length  
8292 8900: contig of 609 bp in length  
gap of unknown length  
8901 9555: contig of 855 bp in length  
gap of unknown length  
9556 10191: contig of 636 bp in length  
gap of unknown length  
10192 10905: contig of 714 bp in length  
gap of unknown length  
10906 11605: contig of 700 bp in length  
gap of unknown length  
11606 12318: contig of 713 bp in length  
gap of unknown length  
12319 12938: contig of 620 bp in length  
gap of unknown length  
12939 13608: contig of 670 bp in length  
gap of unknown length  
13609 14273: contig of 665 bp in length  
gap of unknown length  
14274 14850: contig of 577 bp in length  
gap of unknown length  
14851 15920: contig of 1070 bp in length  
gap of unknown length  
15921 16601: contig of 681 bp in length  
gap of unknown length  
16602 17878: contig of 1277 bp in length  
gap of unknown length  
17879 18931: contig of 1053 bp in length  
gap of unknown length  
18932 19586: contig of 655 bp in length  
gap of unknown length  
19587 20254: contig of 668 bp in length  
gap of unknown length  
20255 21173: contig of 919 bp in length  
gap of unknown length  
21174 21856: contig of 683 bp in length  
gap of unknown length  
21857 22476: contig of 620 bp in length  
gap of unknown length  
22477 23307: contig of 831 bp in length  
gap of unknown length  
23308 24000: contig of 693 bp in length  
gap of unknown length  
24001 24728: contig of 728 bp in length  
gap of unknown length  
24729 25396: contig of 668 bp in length  
gap of unknown length  
25397 26092: contig of 696 bp in length  
gap of unknown length  
26093 26761: contig of 669 bp in length  
gap of unknown length  
26762 28159: contig of 1398 bp in length  
gap of unknown length  
28160 28839: contig of 680 bp in length  
gap of unknown length  
28840 29509: contig of 670 bp in length  
gap of unknown length  
29510 30209: contig of 700 bp in length  
gap of unknown length  
30210 30932: contig of 723 bp in length  
gap of unknown length  
30933 31588: contig of 656 bp in length  
gap of unknown length  
31589 32225: contig of 637 bp in length  
gap of unknown length  
32226 32854: contig of 629 bp in length  
gap of unknown length  
32855 33476: contig of 622 bp in length  
gap of unknown length  
33477 34076: contig of 600 bp in length  
gap of unknown length  
34077 34812: contig of 736 bp in length  
gap of unknown length  
34813 35457: contig of 645 bp in length  
gap of unknown length  
35458 36131: contig of 674 bp in length  
gap of unknown length  
36132 36774: contig of 643 bp in length  
gap of unknown length  
36775 37369: contig of 595 bp in length  
gap of unknown length  
37370 38026: contig of 657 bp in length  
gap of unknown length  
38027 38662: contig of 636 bp in length  
gap of unknown length  
38663 39286: contig of 624 bp in length  
gap of unknown length  
39287 39370: contig of 84 bp in length  
gap of unknown length  
39371 40404: contig of 1034 bp in length  
gap of unknown length  
40405 41437: contig of 1033 bp in length  
gap of unknown length  
41438 42318: contig of 881 bp in length  
gap of unknown length  
42319 42486: contig of 168 bp in length  
gap of unknown length

\* 42487 43151: contig of 665 bp in length  
 \* gap of unknown length  
 \* 43152 44251: contig of 1100 bp in length  
 \* gap of unknown length  
 \* 44252 45291: contig of 1040 bp in length  
 \* gap of unknown length  
 \* 45292 46221: contig of 930 bp in length  
 \* gap of unknown length  
 \* 46222 47473: contig of 1252 bp in length  
 \* gap of unknown length  
 \* 47474 48805: contig of 1332 bp in length  
 \* gap of unknown length  
 \* 48806 49527: contig of 722 bp in length  
 \* gap of unknown length  
 \* 49528 50567: contig of 1040 bp in length  
 \* gap of unknown length  
 \* 50568 52219: contig of 1652 bp in length  
 \* gap of unknown length  
 \* 52220 53518: contig of 1299 bp in length  
 \* gap of unknown length  
 \* 53519 53635: contig of 117 bp in length  
 \* gap of unknown length  
 \* 53636 54472: contig of 837 bp in length  
 \* gap of unknown length  
 \* 54473 55286: contig of 814 bp in length  
 \* gap of unknown length  
 \* 55287 56876: contig of 1590 bp in length  
 \* gap of unknown length  
 \* 56877 57989: contig of 1113 bp in length  
 \* gap of unknown length  
 \* 57990 59011: contig of 1022 bp in length  
 \* gap of unknown length  
 \* 59012 60361: contig of 1350 bp in length  
 \* gap of unknown length  
 \* 60362 61992: contig of 1831 bp in length  
 \* gap of unknown length  
 \* 61993 63418: contig of 1426 bp in length  
 \* gap of unknown length  
 \* 63419 64754: contig of 1336 bp in length  
 \* gap of unknown length  
 \* 64755 65992: contig of 1238 bp in length  
 \* gap of unknown length  
 \* 65993 67163: contig of 1171 bp in length  
 \* gap of unknown length  
 \* 67164 68719: contig of 1556 bp in length  
 \* gap of unknown length  
 \* 68720 70390: contig of 1671 bp in length.

## FEATURES

Source  
 1. .70390  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"

Query Match 9.1%; Score 18; DB 41; Length 70390;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TAGAAAAATATATATT 107  
 |||||  
 Db 58294 TAGAAAAATATATATT 58277

## RESULT 43

AB024028 70952 bp DNA PLN 20-NOV-1999  
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone: K1G2,  
 complete sequence.  
 AB024028  
 AB024028.1 GI:4519187  
 ACCESSION  
 VERSION  
 KEYWORDS HTG.  
 SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone\_lib:Mitsui P1  
 Clone:K1G2.  
 ORGANISM Arabidopsis thaliana

## REFERENCE

AUTHORS Nakamura,Y.  
 TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. I  
 JOURNAL Unpublished (1998)  
 REFERENCE 2 (bases 1 to 70952)  
 AUTHORS Nakamura,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-FEB-1999) to the DDBJ/EMBL/GenBank databases.

Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of  
 Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan  
 (E-mail:ynakam@kazusa.or.jp, Tel:+81-438-52-3935,  
 Fax:+81-438-52-3934)

## FEATURES

Location/Qualifiers  
 source  
 1. .70952

/organism="Arabidopsis thaliana"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /chromosome="3"  
 /clone\_lib="Mitsui P1"

BASE COUNT 23075 a 12412 c 12631 g 22834 t  
 ORIGIN

Query Match 9.1%; Score 18; DB 7; Length 70952;

Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 91 AGAAAAATATATATTG 108  
 |||||  
 Db 16661 AGAAAAATATATATTG 16678

## RESULT 44

AC016217 71293 bp DNA HTG 23-NOV-1999  
 LOCUS Homo sapiens clone RP11-26F18, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION  
 ACCESSION AC016217  
 VERSION AC016217.1 GI:6466577  
 KEYWORDS HTG: HTGS\_PHASE0.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## REFERENCE

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhgaiter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagois,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczy,J., Lieu,C., Locke,C., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stenger-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RN/RepeatMasker.html

## TITLE

## JOURNAL

## COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIGR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)  
----- Project Information  
Center project name: I4628  
Center clone name: 26\_F\_18  
-----

\* NOTE: This record contains 81 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely for  
\* arbitrary. low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

\* 1  
\* 845: contig of 845 bp in length  
\* gap of unknown length  
\* 846 1716: contig of 871 bp in length  
\* gap of unknown length  
\* 1717 2589: contig of 873 bp in length  
\* gap of unknown length  
\* 2590 3446: contig of 857 bp in length  
\* gap of unknown length  
\* 3447 4294: contig of 848 bp in length  
\* gap of unknown length  
\* 4295 5182: contig of 888 bp in length  
\* gap of unknown length  
\* 5183 6077: contig of 895 bp in length  
\* gap of unknown length  
\* 6078 6965: contig of 888 bp in length  
\* gap of unknown length  
\* 6966 7812: contig of 847 bp in length  
\* gap of unknown length  
\* 7813 8707: contig of 895 bp in length  
\* gap of unknown length  
\* 8708 9587: contig of 880 bp in length  
\* gap of unknown length  
\* 9588 10407: contig of 820 bp in length  
\* gap of unknown length  
\* 10408 11312: contig of 905 bp in length  
\* gap of unknown length  
\* 11313 12178: contig of 866 bp in length  
\* gap of unknown length  
\* 12179 13063: contig of 885 bp in length  
\* gap of unknown length  
\* 13064 13929: contig of 866 bp in length  
\* gap of unknown length  
\* 13930 14804: contig of 875 bp in length  
\* gap of unknown length  
\* 14805 15681: contig of 877 bp in length  
\* gap of unknown length  
\* 15682 16549: contig of 868 bp in length  
\* gap of unknown length  
\* 16550 17445: contig of 896 bp in length  
\* gap of unknown length  
\* 17446 18279: contig of 834 bp in length  
\* gap of unknown length  
\* 18280 19157: contig of 878 bp in length  
\* gap of unknown length  
\* 19158 20016: contig of 859 bp in length  
\* gap of unknown length  
\* 20017 20861: contig of 845 bp in length  
\* gap of unknown length  
\* 20862 21726: contig of 865 bp in length  
\* gap of unknown length  
\* 21727 22606: contig of 880 bp in length  
\* gap of unknown length  
\* 22607 23518: contig of 912 bp in length  
\* gap of unknown length

\* 23519 24371: contig of 853 bp in length  
\* gap of unknown length  
\* 24372 25255: contig of 884 bp in length  
\* gap of unknown length  
\* 25256 26130: contig of 875 bp in length  
\* gap of unknown length  
\* 26131 27011: contig of 881 bp in length  
\* gap of unknown length  
\* 27012 27910: contig of 899 bp in length  
\* gap of unknown length  
\* 27911 28780: contig of 870 bp in length  
\* gap of unknown length  
\* 28781 29666: contig of 886 bp in length  
\* gap of unknown length  
\* 29667 30535: contig of 869 bp in length  
\* gap of unknown length  
\* 30536 31418: contig of 883 bp in length  
\* gap of unknown length  
\* 31419 32331: contig of 913 bp in length  
\* gap of unknown length  
\* 32332 33232: contig of 901 bp in length  
\* gap of unknown length  
\* 33233 34112: contig of 880 bp in length  
\* gap of unknown length  
\* 34113 35013: contig of 901 bp in length  
\* gap of unknown length  
\* 35014 35911: contig of 898 bp in length  
\* gap of unknown length  
\* 35912 36814: contig of 903 bp in length  
\* gap of unknown length  
\* 36815 37703: contig of 889 bp in length  
\* gap of unknown length  
\* 37704 38555: contig of 852 bp in length  
\* gap of unknown length  
\* 38556 39457: contig of 902 bp in length  
\* gap of unknown length  
\* 39458 40330: contig of 873 bp in length  
\* gap of unknown length  
\* 40331 41212: contig of 882 bp in length  
\* gap of unknown length  
\* 41213 42096: contig of 884 bp in length  
\* gap of unknown length  
\* 42097 42972: contig of 876 bp in length  
\* gap of unknown length  
\* 42973 43868: contig of 896 bp in length  
\* gap of unknown length  
\* 43869 44749: contig of 881 bp in length  
\* gap of unknown length  
\* 44750 45637: contig of 888 bp in length  
\* gap of unknown length  
\* 45638 46504: contig of 867 bp in length  
\* gap of unknown length  
\* 46505 47375: contig of 871 bp in length  
\* gap of unknown length  
\* 47376 48265: contig of 890 bp in length  
\* gap of unknown length  
\* 48266 49145: contig of 880 bp in length  
\* gap of unknown length  
\* 49146 50026: contig of 881 bp in length  
\* gap of unknown length  
\* 50027 50906: contig of 880 bp in length  
\* gap of unknown length  
\* 50907 51796: contig of 890 bp in length  
\* gap of unknown length  
\* 51797 52661: contig of 865 bp in length  
\* gap of unknown length  
\* 52662 53549: contig of 888 bp in length  
\* gap of unknown length  
\* 53550 54437: contig of 888 bp in length  
\* gap of unknown length  
\* 54438 55331: contig of 894 bp in length  
\* gap of unknown length

```

* 55332 56211: contig of 880 bp in length
* 56212 57077: contig of 866 bp in length
* 57078 57963: contig of 885 bp in length
* 57964 58866: contig of 903 bp in length
* 58867 59738: contig of 872 bp in length
* 59739 60635: contig of 897 bp in length
* 60636 61519: contig of 884 bp in length
* 61520 62425: contig of 906 bp in length
* 62426 63299: contig of 874 bp in length
* 63300 64178: contig of 879 bp in length
*
Query Match          9.18; Score 18; DB 44; Length 71293;
Best Local Similarity 100.08; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AGCATTTTACAGACAAG 166
|||||
Db 34863 AGCATTTTACAGACAAG 34880

RESULT 45
AF191071/c          DNA          PRI          11-OCT-1999
LOCUS              Homo sapiens chromosome 8 clone BAC 388D06, complete sequence.
DEFINITION
ACCESSION          AF191071
VERSION            AF191071.1 GI:6018318
KEYWORDS            HTG.
SOURCE              human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS            Reichwald,K., Baumgart,C., Dette,M., Menzel,U., Schillhabel,M.,
Weng,G. and Rosenthal,A.
TITLE              Direct Submission
JOURNAL            Submitted (29-SEP-1999) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..88481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="BAC 388D06"
repeat_region      1..72
/rpt_family="LIP3_5end"
repeat_region      68..1499
/rpt_family="L1"
exon               286..563
/note="GRAIL"
evidence=not_experimental
repeat_region      1353..2244 "LIP8"
/rpt_family="LIP8"
exon               1851..2027
/note="GRAIL"
evidence=not_experimental
repeat_region      2256..2564
/rpt_family="AluSp"
repeat_region      2789..2811
/note="AT-rich"
evidence=not_experimental
repeat_region      2880..2932
/rpt_family="Low_complexity"
evidence=not_experimental
/note="AT-rich"
/rpt_family="Low_complexity"

```

---

```

repeat_region      3366..3388
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
exon               4271..4750
/note="MZF"
evidence=not_experimental
complement(4378..4937)
/note="GENSCAN"
complement(4397..4422)
/note="XPOUND"
evidence=not_experimental
4468..4490
/note="XPOUND"
evidence=not_experimental
4492..4514
/note="XPOUND"
evidence=not_experimental
complement(5171..5319)
/rpt_family="MIR"
5471..5638
/note="MZF"
evidence=not_experimental
complement(5859..5901)
/note="AT-rich"
/rpt_family="Low_complexity"
5975..6002
/note="AT-rich"
/rpt_family="Low_complexity"
complement(7160..7447)
/rpt_family="AluX"
complement(7891..7933)
/note="(TAGA)n"
/rpt_family="Simple_repeat"
8060..8121
/rpt_family="MLT1"
complement(8823..8905)
/note="MZF"
evidence=not_experimental
9007..9100
/note="AT-rich"
/rpt_family="Low_complexity"
complement(9632..10675)
/rpt_family="LIMC1"
10202..10334
/note="MZF"
evidence=not_experimental
complement(10543..13163)
/rpt_family="LIM4_orf2"
complement(13170..13373)
/rpt_family="MLT2B"
complement(13319..13358)
/note="XPOUND"
evidence=not_experimental
13374..14061
/rpt_family="LIME1"
14070..14283
/rpt_family="LIMEc_5end"
14177..14234
/note="MZF"
evidence=not_experimental
15313..15404
/note="GRAIL"
evidence=not_experimental
15313..15404
/note="MZF"
evidence=not_experimental
complement(16097..16540)
/rpt_family="LIMEa_5end"
16560..18253
/rpt_family="LIM2_orf2"
complement(17617..17816)
/rpt_family="MER30"
complement(17635..17698)

```

```

/repeat_region /note="MZF"
/evidence-not_experimental
complement(18235..20096)
/exon /rpt_family="Tigger1"
complement(18382..18510)
/exon /note="GRAIL"
/evidence-not_experimental
complement(19602..19747)
/exon /note="GRAIL"
/evidence-not_experimental
complement(19602..19747)
/exon /note="GENSCAN"
complement(20169..20407)
/repeat_region /rpt_family="MIR"
complement(20575..20701)
/exon /note="GRAIL"
/evidence-not_experimental
22106..22399
/repeat_region /rpt_family="Alusx"
complement(22402..22449)
/repeat_region /note="(CAAT)n"
complement(22448..22503)
/repeat_region /rpt_family="Simple_repeat"
complement(22994..23212)
/repeat_region /rpt_family="LIMA3"
complement(23501..23948)
/repeat_region /rpt_family="LIME1"
23969..23989
/repeat_region /note="AT_rich"
complement(24825..25242)
/repeat_region /rpt_family="LIMA6"
complement(26521..26547)
/repeat_region /note="AT_rich"
complement(26521..26547)
/repeat_region /rpt_family="Low_complexity"
26816..26850
/repeat_region /note="AT_rich"
complement(27256..27855)
/repeat_region /rpt_family="LIMA2"
27952..28079
/repeat_region /rpt_family="LIMD_Send"
28098..28445
/repeat_region /rpt_family="LIMD_Send"
28471..28643
/repeat_region /rpt_family="LIM4_orf2"
28663..28846
/repeat_region /rpt_family="LIM4_orf2"
28857..29228
/repeat_region /rpt_family="LIM4_orf2"
complement(29236..29746)
/repeat_region /rpt_family="MER74A"
complement(29311..29369)
/exon /note="GENSCAN"
29761..30627
/repeat_region /rpt_family="LIMD1"
29874..29921
/repeat_region /note="(TA)n"
complement(30468..30584)
/exon /note="GRAIL"
complement(30642..30684)
/repeat_region /note="AT_rich"
complement(30706..30755)
/repeat_region /rpt_family="Low_complexity"
complement(30843..30927)
/exon /note="GRAIL"
/evidence-not_experimental

```

```

/repeat_region 30979..31064
/note="(TA)n"
complement(30980..31031)
/repeat_region /note="(CA)n"
complement(31080..31211)
/exon /rpt_family="Simple_repeat"
complement(31080..31211)
/note="GRAIL"
/evidence-not_experimental
complement(31086..31194)
/exon /note="MZF"
/evidence-not_experimental
complement(31086..31211)

```

Query Match 9.1%; Score 18; DB 40; Length 88481;  
 Best Local Similarity 100.0%; Pred.No. 39;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TTAGGGAATTTTGGTTG 74  
 |||||

Db 17960 TTAGGGAATTTTGGTTG 17943

Search completed: April 6, 2000, 21:16:15  
 Job time: 50360 sec



---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 6, 2000, 21:16:15 ; Search time 9481.22 Seconds  
(without alignments)  
-43.875 Million cell updates/sec

Title: US-09-090-672B-11  
Perfect score: 137  
Sequence: 1 CCACCGACCTGCTGATGC.....TCGTTTCAGTCCTGGTCTCT 137

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0  
Searched: 821193 seqs, -1518192014 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pil.\*  
8: gb\_pl2.\*  
9: gb\_pr1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_sts.\*  
14: gb\_sy.\*  
15: gb\_un.\*  
16: gb\_vi.\*  
17: em\_fun.\*  
18: em\_hum1.\*  
19: em\_hum2.\*  
20: em\_in.\*  
21: em\_on.\*  
22: em\_or.\*  
23: em\_ov.\*  
24: em\_pat.\*  
25: em\_ph.\*  
26: em\_pl.\*  
27: em\_ro.\*  
28: em\_sts.\*  
29: em\_sy.\*  
30: em\_un.\*  
31: em\_vi.\*  
32: gb\_htg1.\*  
33: gb\_htg2.\*  
34: gb\_in1.\*  
35: gb\_in2.\*  
36: em\_ba1.\*  
37: em\_ba2.\*  
38: em\_hum3.\*  
39: em\_hum4.\*  
40: gb\_pr4.\*  
41: gb\_htg3.\*  
42: gb\_htg4.\*  
43: gb\_htg5.\*

44: gb\_htg6.\*  
45: gb\_htg7.\*  
46: em\_htg1.\*  
47: em\_htg2.\*  
48: em\_htg3.\*  
49: em\_hum5.\*  
50: gb\_pl3.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	137	100.0	150228	40	AC003071	AC003071 Homo sapi
C 2	20	14.6	22287	45	AC017236	AC017236 Drosophill
C 3	20	14.6	78154	10	AC002466	AC002466 Human BAC
C 4	20	14.6	89743	12	AC003062	AC003062 Mouse Chr
C 5	20	14.6	149625	42	AC010025	AC010025 Drosophill
C 6	20	14.6	174063	41	AC010143	AC010143 Homo sapi
C 7	20	14.6	177591	33	AC006563	AC006563 Drosophill
C 8	20	14.6	180513	44	AC008020	AC008020 Mus muscu
C 9	20	14.6	189870	42	AC010104	AC010104 Homo sapi
C 10	19	13.9	7080	9	AB024745	AB024745 Homo sapi
C 11	19	13.9	50511	11	AC005214	AC005214 Homo sapi
C 12	19	13.9	86329	11	HS403M6	AL033392 Human DNA
C 13	19	13.9	135855	10	HS135E14	AJ010598 Homo sapi
C 14	19	13.9	155450	40	AC005951	AC005951 Homo sapi
C 15	19	13.9	157176	40	AC007092	AC007092 Homo sapi
C 16	19	13.9	160798	44	AC012536	AC012536 Homo sapi
C 17	19	13.9	172048	10	HS179N16	Z95152 Homo sapien
C 18	19	13.9	204027	42	AC007158	AC007158 Homo sapi
C 19	18	13.1	270	9	HUMVIPHM1	M14618 Human vasoa
C 20	18	13.1	474	13	G24450	G24450 human ST5 W
C 21	18	13.1	5190	11	HSU66615	U66615 Human SWI/S
C 22	18	13.1	14837	41	AC008821	AC008821 Homo sapi
C 23	18	13.1	23536	11	AC002120	AC002120 Homo sapi
C 24	18	13.1	33602	10	U73643	U73643 Human Chrom
C 25	18	13.1	35967	43	AC014889	AC014889 Drosophill
C 26	18	13.1	37027	11	HS506G2B	Z82976 Human DNA s
C 27	18	13.1	44294	10	AC000394	AC000394 Genomic s
C 28	18	13.1	48925	42	AC012318	AC012318 Homo sapi
C 29	18	13.1	55707	41	AC011108	AC011108 Homo sapi
C 30	18	13.1	81542	7	AB026647	AB026647 Arabidops
C 31	18	13.1	81874	11	HS931E15	AL023575 Human DNA
C 32	18	13.1	86765	11	AC004472	AC004472 Homo sapi
C 33	18	13.1	89328	11	HS398C22	Z93784 Homo sapien
C 34	18	13.1	89818	11	AC002126	AC002126 Homo sapi
C 35	18	13.1	93841	41	AC009012	AC009012 Homo sapi
C 36	18	13.1	95425	32	HSJ508G21	Z82213 Homo sapien
C 37	18	13.1	96099	41	AC011336	AC011336 Homo sapi
C 38	18	13.1	100091	44	AC013370	AC013370 Homo sapi
C 39	18	13.1	100997	11	HS262D12	Z92927 Homo sapien
C 40	18	13.1	109210	11	HSJ863C7	AL049761 Human DNA
C 41	18	13.1	109891	10	HS112K5	Z85987 Human DNA s
C 42	18	13.1	110000	41	AC010679_2	Continuation (3 of
C 43	18	13.1	110000	43	AC015845_0	AC015845 Homo sapi
C 44	18	13.1	110000	43	AC015845_2	Continuation (3 of
C 45	18	13.1	121345	43	AC013673	AC013673 Homo sapi

# ALIGNMENTS

RESULT 1  
AC003071/c  
LOCUS  
DEFINITION  
AC003071  
AC003071 150228 bp DNA  
Homo sapiens BAC clone BK085E05 from 22q12.1-qter, complete  
sequence.  
PRI 18-MAR-1999

VERSION	AC003071.1	GI:2598643	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 150228)		
AUTHORS	Sulston,J.E. and Waterston,R.		
TITLE	Toward a complete human genome sequence		
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)		
MEDLINE	95063792		
REFERENCE	2 (bases 1 to 150228)		
AUTHORS	Bradshaw,H., Graves,T., Sutterer,C. and Ozersky,P.		
TITLE	The sequence of Homo sapiens BAC clone BK085E05		
JOURNAL	Unpublished (1999)		
REFERENCE	3 (bases 1 to 150228)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-NOV-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
	4 (bases 1 to 150228)		
AUTHORS	Waterston,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-MAR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
COMMENT	SUBMITTED BY: WUGSC Genome Sequencing Center Department of Genetics Washington University St. Louis MO 63108, USA http://genome.wustl.edu/gsc mailto:sapiens@watson.wustl.edu		
	NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.		
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.		
	MAPPING INFORMATION: This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the chromosome 22 mapping group at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22/		
	SOURCE INFORMATION: This clone is from the human BAC library described by U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).		
	VECTOR: pBAC108L Selection: chloramphenicol		
FEATURES	NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is DJ0394A18, 200 bp overlap; the clone sequenced to the right is DJ43804, 200 bp overlap. Actual start of this clone is at base position 1 of BK085E05.		
source	Location/Qualifiers 1..150228 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="22" /map="22q12.1-qter" /clone_lib="BKA" /clone="BK085E05" <19..17848 /gene="WUGSC:H_BK085E05.1"		
gene			

```

/rpt_family="Alu"
10969..11269
/rpt_family="Alu"
12297..12773
/genes="WUGSC:H_BK085E05.1"
/notes="match to EST R60817 (NID:g831512) yh04b10.r1"
12430..12694
/genes="WUGSC:H_BK085E05.1"
/notes="match to EST T19045 (NID:g601088)"
12821..12935
/genes="WUGSC:H_BK085E05.1"
/notes="similar to EST AA230681 (NID:g1852980) mv76a09.r1"
14855..14745
/genes="WUGSC:H_BK085E05.1"
/notes="similar to EST AA230681 (NID:g1852980) mv76a09.r1"
complement(14682..14745)
/genes="WUGSC:H_BK085E05.1"
/notes="match to EST AA508154 (NID:g2244593) ng93g05.s1"
15909..15936
/rpt_family="POLY_A"
16061..16367
/rpt_family="Alu"
complement(17675..18023)
/notes="match to EST AA508154 (NID:g2244593) ng93g05.s1"
17875..17821
/genes="WUGSC:H_BK085E05.1"
/notes="similar to EST AA230681 (NID:g1852980) mv76a09.r1"
17755..18059
/notes="match to EST R24323 (NID:g779211) yg32e07.r1"
18123..18397
/notes="match to EST Z44461 (NID:g573593)"
18177..18531
/notes="match to EST R13425 (NID:g766501) yf76g10.r1"
18453..18915
/notes="match to EST N39988 (NID:g1163533) yx97g05.r1"
18501..18809
/notes="match to EST AA056351 (NID:g1548753) z165g03.s1"
18676..19007
/notes="match to EST W81505 (NID:g1392696) zd88g06.r1"
18779..19138
/notes="match to EST AA370856 (NID:g2023174)"
18784..19060
/notes="match to EST R91889 (NID:g959429) yq04c12.r1"
18842..19231
/notes="match to EST AA099258 (NID:g1645104) zk84e08.r1"
18945..19353
/notes="match to EST AA024874 (NID:g1489999) ze80f03.r1"
19172..19501
/notes="match to EST N26968 (NID:g1141316) yx18f03.r1"
complement(19287..19546)

Query Match          100.0%; Score 137; DB 40; Length 150228;
Best Local Similarity 100.0%; Pred. No. 2.7e-69;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGTGATGCTTTTCTATCTGACTTCTTTCAGAGGACCCCTGAAGACACT 60
Db 16074 CCACCGCACCTGGTGATGCTTTTCTATCTGACTTCTTTCAGAGGACCCCTGAAGACACT 16015

Qy 61 AAGTGGAAATCTTTCCTGAAGCTTCGAAGCTTCAAGCTAAACAATCTCTGGAAGATCACCTCT 120
Db 16014 AAGTGGAAATCTTTCCTGAAGCTTTCCTGAAGCTTCAAGCTAAACAATCTCTGGAAGATCACCTCT 15955

Qy 121 GTTCAGTCTGCTCTCT 137
Db 15954 GTTCAGTCTGCTCTCT 15938

RESULT 2
AC017236/c 22287 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.

```

```

ACCESSION AC017236
VERSION AC017236.1 GI:6553750
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 22287)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10209970 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
1..22287
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 6759 a 4503 c 4289 g 6736 t
ORIGIN
Query Match          14.6%; Score 20; DB 45; Length 22287;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 CTTTCTATCTGACTTCTTT 39
Db 9360 CTTTCTATCTGACTTCTTT 9341

RESULT 3
AC002466
LOCUS Human BAC clone RG354L07 from 7q31, complete sequence.
DEFINITION AC002466
ACCESSION AC002466
VERSION AC002466.1 GI:2337861
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 78154)
AUTHORS Miller,N and Gibson,A.
TITLE The sequence of H. sapiens BAC clone RG354L07
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 78154)
AUTHORS Waterston,R.
JOURNAL Direct Submission
COMMENT Submitted (20-AUG-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by

sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7> or send <mailto:egreen@nhgri.nih.gov>

#### SOURCE INFORMATION:

This clone is from a release of the human BAC library. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11

Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The actual start of this clone is at base position 1 of RG354L07; actual end is at 78154 of RG354L07. The orientation of this clone is unknown.

This clone contains STS's SWSS2071 (NID:gl113356) and SWSS2179 (NID:gl113411).

#### FEATURES

```

    source
        1..78154
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="7"
        /clone="RG354L07"
        /clone.lib="CIITB-HS-A"
        /map="7q31"
        complement(575..827)
        /rpt_family="ALU"
        1985..2017
        /rpt_family="L1"
        3425..3629
        /rpt_family="MER"
        3721..3940
        /rpt_family="MER"
        complement(4181..4594)
        /rpt_family="MER"
        complement(4625..4923)
        /rpt_family="L1"
        complement(4939..5482)
        /rpt_family="L1"
        5991..6283
        /rpt_family="ALU"
        complement(6957..7181)
        /rpt_family="MER"
        complement(7430..7453)
        /rpt_family="L1"
        7957..7992
        /rpt_family="L1"
        10650..10692
        /rpt_family="L1"
        12746..12817
        /rpt_family="L1"
        13582..13599
        /rpt_family="L1"
        14112..14318
        /rpt_family="L1"
        16949..17210
        /note="match to EST N73830 (NID:gl231115) za56c01.s1"
        complement(18244..18641)
        /note="match to EST W00873 (NID:gl272993) za56c01.r1"
        20048..20337
        /rpt_family="ALU"

```

```

repeat_region      20338..20367
/rpt_family="L1"
repeat_region      21588..21600
/rpt_family="L1"
repeat_region      22318..22608
/rpt_family="ALU"
repeat_region      25037..25083
/rpt_family="L1"
complement(28356..28928)
/rpt_family="L1"
repeat_region      33433..33459
/rpt_family="L1"
repeat_region      34370..34460
/rpt_family="L1"
repeat_region      34718..34737
/rpt_family="L1"
repeat_region      complement(35485..35775)
/rpt_family="ALU"
repeat_region      36208..37412
/rpt_family="L1"
complement(40472..40559)
/rpt_family="L1"
repeat_region      40482..40507
/rpt_family="L1"
repeat_region      40560..40854
/rpt_family="ALU"
complement(40874..42045)
/rpt_family="L1"
repeat_region      complement(45725..45762)
/rpt_family="L1"
repeat_region      47435..47710
/rpt_family="ALU"
repeat_region      48328..48357
/rpt_family="L1"
repeat_region      complement(49283..49415)
/rpt_family="L1"
repeat_region      complement(54609..54901)
/rpt_family="ALU"
repeat_region      complement(57040..57069)
/rpt_family="L1"
repeat_region      59551..59571
/rpt_family="L1"
repeat_region      59576..59705
/rpt_family="ALU"
complement(60076..60338)
/rpt_family="ALU"
repeat_region      complement(60918..60945)
/rpt_family="L1"
repeat_region      complement(60954..61254)
/rpt_family="ALU"
complement(62605..62633)
/rpt_family="L1"
repeat_region      63387..63437
/rpt_family="L1"
repeat_region      complement(63476..63766)
/rpt_family="ALU"
repeat_region      67301..67331
/rpt_family="L1"
repeat_region      complement(68442..68462)
/rpt_family="L1"
repeat_region      complement(73916..74206)
/rpt_family="ALU"
repeat_region      complement(76596..76628)
/rpt_family="L1"

```

BASE COUNT 25369 a 13876 c 14122 g 24787 t  
ORIGIN

Query Match 14.6%; Score 20; DB 10; Length 78154;  
Best Local Similarity 100.0%; Pred.No. 0.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 TCCAAGCTAATAACAATTCTC 104

```
Db 75221 TCCAGCTAAACAATTCTC 75240
|||||
AC003062 89743 bp DNA ROD 03-MAY-1999
Mouse Chromosome 16 Region Syntenic to DGCRC BAC Clone b264n1,
complete sequence.
AC003062
AC003062.2 GI:4731672
HTG.
SOURCE house musculus
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 89743)
REFERENCE Unpublished
AUTHORS Chen,F. and Roe,B.A.
2 (bases 1 to 89743)
AUTHORS Galli,N., Baldwin,S., Lund,J., Reeves,R., Gong,W., Chen,F.,
Roe,B.A., Emanuel,B.S., Nayak,S., Mickanin,C., Budarf,M.L. and
Buck,C.A.
TITLE A Region of Mouse Chromosome 16 is Syntenic to the Digeorge,
JOURNAL Velo-Cardio-Facial Syndrome Minimal Critical Region
REMARK The genes were identified by comparing with human genomic and cDNA
sequences and RT-PCR of 12 day post conception mouse embryos total
RNA
3 (bases 1 to 89743)
REFERENCE Roe,B.A. Dr.
AUTHORS Direct Submission
TITLE Submitted (06-NOV-1997) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
4 (bases 1 to 89743)
REFERENCE Roe,B.A. Dr.
AUTHORS Direct Submission
TITLE Submitted (16-SEP-1998) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
5 (bases 1 to 89743)
REFERENCE Roe,B.A. Dr.
AUTHORS Direct Submission
TITLE Submitted (23-SEP-1998) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
6 (bases 1 to 89743)
REFERENCE Roe,B.A. Dr.
AUTHORS Direct Submission
TITLE Submitted (05-NOV-1998) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
7 (bases 1 to 89743)
REFERENCE Roe,B.A. Dr.
AUTHORS Direct Submission
TITLE Submitted (03-MAY-1999) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval Room 208, Norman,
OK 73019, USA
COMMENT On May 3, 1999 this sequence version replaced gi:3845373.
FEATURES
source
1. .89743
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="b264n1"
BASE COUNT 24078 a 18138 c 20153 g 27374 t
ORIGIN
Query Match 14.6%; Score 20; DB 12; Length 89743;
Best Local Similarity 100.0%; Pred. No. 0.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 69 TCTTCTTGAAGTCTTCCA 88
|||||
Db 61998 TCTTCTTGAAGTCTTCCA 61979
RESULT 5
AC010025/c DNA HTG 16-OCT-1999
LOCUS Drosophila melanogaster chromosome 3L/69D4 clone RCI98-20B1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 57 unordered pieces.
ACCESSION AC010025 149625 bp
VERSION AC010025.4 GI:6056162
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 149625)
REFERENCE Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
AUTHORS Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Fernandez,C., Ferraguto,D.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Gorrall,J.H., Gorrall,L.L.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrall,J.H., Gorrall,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondajewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahab,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
TITLE Unpublished
JOURNAL 2 (bases 1 to 149625)
REFERENCE Worley,K.C.
AUTHORS Direct Submission
TITLE Submitted (11-SEP-1999) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 16, 1999 this sequence version replaced gi:5916438.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
757: contig of 757 bp in length
1 758: contig of 1264 bp in length
* 2022: contig of 787 bp in length
* 2809: contig of 762 bp in length
* 3571: contig of 768 bp in length
* 4339: contig of 1163 bp in length
* 5502: contig of 839 bp in length
* 6341: contig of 717 bp in length
* 7058: contig of 965 bp in length
* 8022: contig of 845 bp in length
* 8867: contig of 1427 bp in length
* 10294: contig of 1609 bp in length
* 11904: contig of 1180 bp in length
* 13084: contig of 864 bp in length
* 13948: contig of 1482 bp in length
* 15430: contig of 2381 bp in length
* 17811: contig of 1699 bp in length
* 19510: contig of 1879 bp in length
* 21389: contig of 1775 bp in length
* 21389
```





Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,  
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and  
Rubin,G.M.

# Direct Submission

Submitted (12-FEB-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Aug 2, 1999 this sequence version replaced gi:5630021.

For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bugsfritfly@berkeley.edu](mailto:bugsfritfly@berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 14 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 689: contig of 689 bp in length

\* 690 769: gap of unknown length

\* 770 1401: contig of 632 bp in length

\* 1402 1481: gap of unknown length

\* 1482 2084: contig of 603 bp in length

\* 2085 2164: gap of unknown length

\* 2165 2738: contig of 574 bp in length

\* 2739 2818: gap of unknown length

\* 2819 3763: contig of 945 bp in length

\* 3764 3843: gap of unknown length

\* 3844 6770: contig of 2927 bp in length

\* 6771 6850: gap of unknown length

\* 6851 10859: contig of 4019 bp in length

\* 10870 10949: gap of unknown length

\* 10950 24667: contig of 13718 bp in length

\* 24668 24747: gap of unknown length

\* 24748 41201: contig of 16454 bp in length

\* 41202 41281: gap of unknown length

\* 41282 68640: contig of 27359 bp in length

\* 68641 68720: gap of unknown length

\* 68721 95333: contig of 26613 bp in length

\* 95334 95413: gap of unknown length

\* 95414 125111: contig of 29698 bp in length

\* 125112 125191: gap of unknown length

\* 125192 177301: contig of 52110 bp in length

\* 177302 177381: gap of unknown length

\* 177382 177591: contig of 210 bp in length.

\* Location/Qualifiers

\* 1. .177591

\* /organism="Drosophila melanogaster"

\* /strain="v: cn bw sp"

\* /db\_xref="taxon:7227"

\* /chromosome="3"

\* /map="69F1-69F2"

\* /clone="BACR48A05 (D477) RPCI-98 48.A.5"

\* /clone.lib="RPCI-98 (Roswell Park Cancer Institute

\* Drosophila melanogaster BAC library, partial EcoRI in

\* PBACE3.6"

\* BASE COUNT 52325 a 35740 c 35243 g 53234 t 1049 others

## ORIGIN

Query Match

Best Local Similarity 14.6%; Score 20; DB 33; Length 177591;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 20 CTTTCTATCTGACTTCTTT 39

|||||

Db 22042 CTTTCTATCTGACTTCTTT 22023

## RESULT 8

AC008020

## LOCUS

AC008020 180513 bp DNA HTG 23-NOV-1999

Mus musculus, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 10 unordered pieces.

## ACCESSION

AC008020

AC008020.22 GI:6456790

HTG; HTGS-PHASE1.

## KEYWORDS

house mouse.

## SOURCE

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 180513)

Deschamps,S., Oomen,S., Draber,R., Becat,C. and Roe,B.A.

Mus musculus Chromosome 16 PAC Clone p598X13

Unpublished

2 (bases 1 to 180513)

Deschamps,S., Oomen,S. and Roe,B.A.

Direct Submission

Submitted (10-JUL-1999) Department of Chemistry And Biochemistry,

The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

## COMMENT

On Nov 21, 1999 this sequence version replaced gi:6403511.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 10 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 2144: contig of 2144 bp in length

\* 2145 2211: gap of unknown length

\* 2212 10395: contig of 8184 bp in length

\* 10396 10462: gap of unknown length

\* 10463 26456: contig of 15994 bp in length

\* 26457 26523: gap of unknown length

\* 26524 32813: contig of 6290 bp in length

\* 32814 32880: gap of unknown length

\* 32881 50735: contig of 17854 bp in length

\* 50736 50802: contig of 22303 bp in length

\* 50803 73104: contig of 22303 bp in length

\* 73105 73171: gap of unknown length

\* 73172 80446: contig of 7275 bp in length

\* 80447 80512: gap of unknown length

\* 80513 111027: contig of 30515 bp in length

\* 111028 111093: gap of unknown length

\* 111094 138342: contig of 27249 bp in length

\* 138343 138408: gap of unknown length

\* 138409 180513: contig of 42105 bp in length.

\* Location/Qualifiers

\* 1. .180513

\* /organism="Mus musculus"

\* /db\_xref="taxon:10090"

\* BASE COUNT 51816 a 36189 c 37396 g 54499 t 613 others

## ORIGIN

Query Match

Best Local Similarity 14.6%; Score 20; DB 44; Length 180513;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCTTCTCTGAAGTCTTCCA 88

|||||

Db 46125 TCTTCTCTGAAGTCTTCCA 46144

## RESULT 9

AC010104

LOCUS

DEFINITION

AC010104

Homo sapiens clone NH0540C18, \*\*\* SEQUENCING IN PROGRESS \*\*\*; 2

unordered pieces.

AC010104

VERSION

AC010104.2 GI:6136424

KEYWORDS

HTG; HTGS-PHASE1.

SOURCE

human.

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 189870)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189870)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-1999) Genome Sequencing Center, Washington
University, School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63106, USA
COMMENT On Oct 27, 1999 this sequence version replaced gi:5870287.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 71164: contig of 71164 bp in length
* 71165 71182: gap of unknown length
* 71183 189870: contig of 118688 bp in length.
FEATURES Location/Qualifiers
source
1..189870
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0540C18"
BASE COUNT 51911 a 40881 c 41550 g 55510 t 18 others
ORIGIN
Query Match 14.6%; Score 20; DB 42; Length 189870;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCAGCTGGCTGATGC 20
|||||
Db 23673 CCACCGCAGCTGGCTGATGC 23692

RESULT 10
AB024745/c
LOCUS AB024745 7080 bp DNA PRI 09-JUN-1999
DEFINITION Homo sapiens gene for Fe65L2, complete cds.
ACCESSION AB024745
VERSION AB024745.1 GI:5019575
KEYWORDS Fe65L2.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (sites)
AUTHORS Tanahashi,H. and Tabira,T.
TITLE Genome structure and chromosomal mapping of the gene for Fe65L2
interacting with Alzheimer's beta-amyloid precursor protein
JOURNAL Biochem. Biophys. Res. Commun. 258 (2), 385-389 (1999)
MEDLINE 99262108
REFERENCE 2 (bases 1 to 7080)
AUTHORS Tanahashi,H.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-1999) to the DBJ/EMBL/GenBank databases. Hiroshi
Tanahashi, National Institute of Neuroscience, Division of
Demyelinating Disease and Aging; 4-1-1 Ogawahigashi, Kodaira, Tokyo
187-8502, Japan (E-mail:tanahashencnaxp@ncnp.go.jp,
Tel:81-42-341-2711(ex:5163), Fax:81-42-346-1747)
FEATURES Location/Qualifiers
source
1..7080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
```

```
misc_feature
1018 /note="transcription start site"
1052 /note="transcription start site"
gene 1105..6764
/feature="Fe65L2"
join(1105..1153,1435..1598,1680..1756,2626..2886,.2886,.2886,.2996,3122..3249,3500..3505,3648..3762,3900..3984,4248..4331,4595..4710,4845..5036,6528..6764)
/feature="Fe65L2"
/function="interacting with Alzheimer's beta-amyloid precursor protein"
/codon_start=1
/product="Fe65L2"
/protein_id="BAA78674.1"
/db_xref="GI:5019576"
/translation="MLCKDYMLATILVNCDDDLGWDHSLVEAGLPGWRIHDAAGT
YRHVPSGSGTQWRPTWELGDAEDPGTGEGLRPPKGRSFSSLESSLDRSNLSW
YGESYIQSMPEGAKFVRSGLWVEPEDLAPKSSIAVNNCIQQLAOTRSKQPP
DGAWGEGONMLMKDAMSLNPLDHLHLCQPLVHVRVGVGSGKGRDRFAFAS
DKDSCKLCHVFRCDVPAKAIASALHGLCAQILSERVEVSGDASCCSPDIPSDPLR
QVELLDVAQAAQKYEALYMGTLPTVKAMGMDVLENAIGTLTARGDNAWPTMLSVS
DSLMTAHPIOAEASTEPEPLKQCPVRLVTFIGVGRDPTTGLIADLGROFQCAAFWC
OPHAGGLSEAVQACVMYQKCLVASAARGKANGAQARARLRLKRTSMSPSGPLPL
PLKGGVGGAGATPRKRGVFSFLDAERKPSLLHP"
polyA_signal 7058..7063
BASE COUNT 1464 a 1789 c 2017 g 1810 t
ORIGIN
Query Match 13.9%; Score 19; DB 9; Length 7080;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 GATCACCTCTGTTTCAGTCC 129
|||||
Db 2763 GATCACCTCTGTTTCAGTCC 2745

RESULT 11
AC005214/c
LOCUS AC005214 50511 bp DNA PRI 01-JUL-1998
DEFINITION Homo sapiens chromosome 5, BAC clone 319C17 (LBNL H159), complete
sequence.
ACCESSION AC005214
VERSION AC005214.1 GI:3282166
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 50511)
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
MEDLINE 99262108
REFERENCE 2 (bases 1 to 50511)
AUTHORS Ricke,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 50511)
AUTHORS Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES Location/Qualifiers
```



```

misc_feature      complement(27838..27936)
                  /note="GRAIL 2 excellent exon, frame 0"
prim_transcript   join(29553..29612,29893..30060,30138..30215,31580..31709,
32097..32224,32706..32798,33054..33177,33302..33498,
35102..35202)
                  /standard_name="Fe65L2"
                  /note="80%-91% identity rat Y13413 transcriptional
activator FE65"
repeat_region     complement(30471..30874)
                  /rpt_family="Alu"
misc_feature      31341..31456
                  /note="GRAIL 2 excellent exon, frame 1"
misc_feature      31582..31709
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      32108..32222
                  /note="GRAIL 2 excellent exon, frame 2"
misc_feature      32684..32791
                  /note="GRAIL 2 excellent exon, frame 0"
misc_feature      33055..33206
                  /note="GRAIL 2 excellent exon, frame 2"
repeat_region     34605..34751
                  /rpt_family="MIR"

Query Match      13.9%; Score 19; DB 11; Length 50511;
Best Local Similarity 100.0%; Pred. NO.1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 GATCACCCTGTTTCAGTCC 129
      |||||
Db 31223 GATCACCCTGTTTCAGTCC 31205

RESULT 12
HS403M6/c
LOCUS
DEFINITION
      HS403M6 86329 bp DNA PRI 23-NOV-1999
      Human DNA sequence from clone 403M6 on chromosome 6q24.1-25.2.
      Contains two unconnected exons of the gene for Myasthenia Gravis
      autoantigen Gravin, and ESTs, STSS and GSSs, complete sequence.
ACCESSION
      AL033392
VERSION
      AL033392.5 GI:4826487
KEYWORDS
      HTG; Gravin; Myasthenia Gravis autoantigen.
SOURCE
      human.
      ORGANISM
      Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
      Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
      1 (bases 1 to 86329)
      AUTHORS
      Williams, S.
      TITLE
      Direct Submission
      JOURNAL
      Submitted (20-MAY-1999) Sanger Centre, Hinxton, Cambridgeshire,
      CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
      requests: clonerequest@sanger.ac.uk
      On May 13, 1999 this sequence version replaced gi:4578468.
      During sequence assembly data is compared from overlapping clones.
      Where differences are found these are annotated as variations.
      Together with a note of the overlapping clone name. Note that the
      variation annotation may not be found in the sequence submission
      corresponding to the overlapping clone, as we submit sequences with
      only a small overlap as described above.
      The following abbreviations are used to associate primary accession
      numbers given in the feature table with their source databases:
      Eni., EMBL; Swi., SWISSPROT; Tri., TREMBL
      This sequence is the entire insert of clone 403M6. This sequence
      has been finished according to sequence map criteria as follows. An
      attempt is made to resolve all sequencing problems, such as
      compressions and repeats, but not necessarily within known
      annotated human repeat sequence elements (e.g. Alu). Where the
      sequence is ambiguous, there is an annotation using the 'unsure'
      feature key.
      This sequence was generated from part of bacterial clone contigs of
      human chromosome 6, constructed by the Sanger Centre Chromosome 6
      Mapping Group. Further information can be found at
      http://www.sanger.ac.uk/HGP/Chr6
      403M6 is from the library RPC13 constructed at the Roswell Park

```

```

Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCIPAC2.
FEATURES
      Location/Qualifiers
      1..86329
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="6"
      /map="q24.1-25.2"
      /clone="RP3-403M6"
      /clone_lib="RPC1-3"
      201..494
      /note="AluSp repeat: matches 1..299 of consensus"
      550..832
      /note="AluJb repeat: matches 1..300 of consensus"
      1672..1982
      /note="AluSx repeat: matches 1..312 of consensus"
      2091..2330
      /note="L1PA16 repeat: matches 5943..6157 of consensus"
      2331..2630
      /note="AluSg repeat: matches 1..300 of consensus"
      2631..2763
      /note="L1PA16 repeat: matches 5804..5943 of consensus"
      3673..3960
      /note="AluSx repeat: matches 1..305 of consensus"
      3961..3982
      /note="11 copies 2 mer tg 100% conserved"
      4022..4061
      /note="20 copies 2 mer at 80% conserved"
      4097..4379
      /note="AluSx repeat: matches 1..282 of consensus"
      4434..4545
      /note="56 copies 2 mer at 83% conserved"
      4546..4579
      /note="17 copies 2 mer ac 100% conserved"
      4505..4910
      /note="AluJb repeat: matches 1..307 of consensus"
      5749..5994
      /note="AluSx repeat: matches 5..251 of consensus"
      6140..6432
      /note="AluSx repeat: matches 1..305 of consensus"
      6443..6748
      /note="AluSx repeat: matches 1..297 of consensus"
      6780..6877
      /note="AluSp/q repeat: matches 215..312 of consensus"
      7006..7288
      /note="AluSp repeat: matches 1..291 of consensus"
      7289..7765
      /note="match: EST T71187"
      7870..8149
      /note="match: GSS B89849"
      8079..8366
      /note="AluJo repeat: matches 3..294 of consensus"
      8507..8912
      /note="AluSg repeat: matches 1..313 of consensus"
      8913..8994
      /note="L2 repeat: matches 1619..1698 of consensus"
      complement(8992..9405)
      /note="match: STS G27817"
      complement(8995..9731)
      /note="match: ESTs T72883 N58206"
      9427..9722
      /note="AluSx repeat: matches 5..300 of consensus"
      9904..10199
      /note="AluSx repeat: matches 1..297 of consensus"
      10266..10343
      /note="L1MB7 repeat: matches 6099..6173 of consensus"
      10456..10841
      /note="match: GSS AQ134006"
      11936..12074
      /note="AluY repeat: matches 167..307 of consensus"
      13166..13471
      /note="AluSp repeat: matches 2..303 of consensus"
      13662..13977

```

```
repeat_region /note="AluSg repeat: matches 1. .307 of consensus"
13978. .14013
repeat_region /note="18 copies 2 mer tt 97% conserved"
14751. .15089
repeat_region /note="L1NB8 repeat: matches 5840. .6173 of consensus"
15090. .15391
repeat_region /note="AluSp repeat: matches 1. .303 of consensus"
15392. .15828
misc_feature /note="L1NB8 repeat: matches 5386. .5840 of consensus"
15461. .15904
repeat_region /note="match: EST M62284"
16184. .16358
repeat_region /note="L1MC4 repeat: matches 7705. .7880 of consensus"
16360. .16659
repeat_region /note="AluX repeat: matches 1. .298 of consensus"
16689. .16993
repeat_region /note="AluX repeat: matches 1. .304 of consensus"
17093. .17402
repeat_region /note="AluJo repeat: matches 1. .306 of consensus"
complement(17391. .17672)
misc_feature /note="match: STS G08536 G09147"
17528. .17569
repeat_region /note="21 copies 2 mer ga 76% conserved"
17572. .17865
repeat_region /note="AluSx repeat: matches 1. .282 of consensus"
17866. .17957
repeat_region /note="L1MC4 repeat: matches 7413. .7505 of consensus"
17958. .18240
repeat_region /note="AluYb8 repeat: matches 1. .298 of consensus"
18241. .18284
repeat_region /note="L1MC4 repeat: matches 7370. .7413 of consensus"
18786. .19083
repeat_region /note="AluSg repeat: matches 5. .301 of consensus"
19108. .19412
repeat_region /note="AluYb repeat: matches 1. .298 of consensus"
19857. .20147
repeat_region /note="AluYb repeat: matches 3. .292 of consensus"
22576. .22624
repeat_region /note="L2 repeat: matches 2701. .2749 of consensus"
22979. .23014
repeat_region /note="18 copies 2 mer tg 94% conserved"
23238. .23549
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
23708. .23999
repeat_region /note="AluY repeat: matches 2. .293 of consensus"
24736. .24898
repeat_region /note="FRAM repeat: matches 5. .161 of consensus"
25209. .25316
repeat_region /note="L1ME repeat: matches 5683. .5790 of consensus"
25549. .25646
repeat_region /note="U6 repeat: matches 1. .107 of consensus"
25669. .25839
repeat_region /note="FRAM repeat: matches 0. .175 of consensus"
26724. .27021
repeat_region /note="AluJo repeat: matches 11. .305 of consensus"
27122. .27443
repeat_region /note="AluSx repeat: matches 1. .296 of consensus"
27629. .27925
misc_feature /note="match: GSS B75668"
27677. .27683
misc_feature /note="IS150 excised."
28611. .28677
repeat_region /note="MIR repeat: matches 97. .166 of consensus"
29085. .29243
repeat_region /note="L1ME3A repeat: matches 5947. .6118 of consensus"
29378. .29689
repeat_region /note="AluYb repeat: matches 1. .307 of consensus"
30419. .30452
repeat_region /note="17 copies 2 mer tg 100% conserved"
30914. .31194
repeat_region /note="AluSg repeat: matches 18. .298 of consensus"
31963. .32272
repeat_region /note="AluY repeat: matches 1. .309 of consensus"

repeat_region 32275. .32594
/note="AluYb repeat: matches 1. .312 of consensus"
32627. .33244
repeat_region /note="L1PA7 repeat: matches 5520. .6138 of consensus"
33267. .33567
repeat_region /note="AluSg repeat: matches 1. .301 of consensus"
complement(34525. .35194)
misc_feature /note="match: GSS AQ268889"
complement(34670. .35222)
misc_feature /note="match: GSS AQ353171"
35217. .35639
misc_feature /note="match: GSS AQ200990 AQ505480"
35840. .35943
repeat_region /note="AluSp repeat: matches 1. .300 of consensus"
36195. .36288

Query Match 13.9% Score 19: DB 11: Length 86329;
Best Local Similarity 100.0%: Pred No. 1.5;
Matches 19: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 1 CCACCGCACCTGGCTGATG 19
|||||
Db 41543 CCACCGCACCTGGCTGATG 41525

RESULT 13
HS135E14
LOCUS HS135E14 135855 bp DNA PRI 03-NOV-1999
DEFINITION Homo sapiens chromosome 21 PAC RPCIP704E14135Q2, complete sequence.
ACCESSION AJ010598
VERSION AJ010598.1 GI:3559851
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 135855)
AUTHORS Brandt,P., Dose,S., Grimm,M., Groet,J., Hornischer,K.,
Loehner,T.H., Nizetic,D., Scharfe,M., Schoen,O., Yaspo,M.L. and
Bioecker,H.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-1998) GBF, Dept. of Genome Analysis, Mascheroder
Weg 1, D-38124 Braunschweig, Germany, E-mail: hornischer@gbf.de,
bioecker@gbf.de
COMMENT Sequence overlaps with Acc.Nr. AJ010597
Collaborators:
Center for applied Molecular Biology
School of Pharmacy, University of London
29-39 Brunswick Square, London, WC1N 1AX, UK
e-mail: j.groet@chemb.ulsop.ac.uk
and
GBF, Dept. of Genome Analysis,
Mascheroder Weg 1, D-38124 Braunschweig, Germany,
and
Max-planck-Institut fuer Molekulare Genetik
Innestrasse 73, D-14195 Berlin-Dahlem, Germany
All annotations in this database entry are developed by
computational tools. It is therefore not explicitly noted in the
feature lines that evidence is not experimental.
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++
+ Analysis and annotation were performed with the automatic
+ 'first-pass' annotation and submission tool
+ 'AnnoMitter' (Hornischer & Bioecker).
+ Programs used by 'AnnoMitter':
+ +++++
> GeneFinder (Green), Vers. 084
> Organism: human
> GenScan (Burge & Karlin), Vers. 1.0
> Used matrix: vertebrate; Minimum score: 0
> Grail (Xu et al.), Vers. 1.3
> Organism: human
> Mzeif (Zhang)
```

```

: Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas
& Skolnick)
: Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats':
BLASTN 2.0.9 (Altschul et al.)
: Database(s): * RepBase (human), released 22-DEC-1995
: * RepBase (primate), released 22-DEC-1995
: * RepBase (mammal), released 22-DEC-1995
: Minimum score: 60;
Minimum identity: 70 %;
: 'ESTs': BLASTN 2.0.9 (Altschul et al.)
: Database(s): * emb1 (EST), Vers. 60 (16-SEP-1999)
: * embnew (EST), Vers. 60+ (20-OCT-1999)
: Using unmasked
sequence
: Minimum score: 60; Minimum identity: 70 %;
: 'GSSs': BLASTN 2.0.9 (Altschul et al.)
: Database(s): * emb1 (GSS), Vers. 60 (16-SEP-1999)
: * embnew (GSS), Vers. 60+ (20-OCT-1999)
: Using unmasked
sequence
: Minimum score: 60; Minimum identity: 70 %;
> 'Tandem Repeats': GDE 2.2 option 'tandem'
: Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
: Treat N's as mismatches? YES; Allow uniform consensi? NO >
> 'Inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'sputnik' (Abajian) > 'CpG
Islands': GDE 2.2 option 'cpg'
: CpG island region size 100 bp;
: Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
e-PCR (Schuler)
: Margin: 50; Number of mismatches allowed: 0; Word size: 7
: STS database: 'dbSTS markers'
> 'LRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
Location/Qualifiers
1..135855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="RPCIP704E14135Q2"
/map="21q11.2"
1..12050
/feature="overlapping bases 93949..105999 in AJ010597"
72..83
/feature="AGC repeat"
complement(415..554)
/feature="match: EST AA482201"
417..554
/feature="GENSCAN prediction, score = 4.94; GRAIL, score =
100%, comment = excellent; MZEF prediction, score = 0.666"
1736..1737
/feature="insertion in AJ010597, pos. 95684..95686"
/replace="ttc"
complement((2237..2345)..(2300..2388))
/feature="match: GSSs AQ095980 AQ171887 AQ387155 AQ139774
AQ356842 AQ355302 B76031 AQ518812 AQ663333 B80647 AQ729504
B53194 B57183 AQ505725 AQ035195 AQ822466 AQ037994 AQ081956
AQ003991 AQ761322 B84086 AQ300632 B95497 AQ744175 AQ100967
AQ231861 B55560 AQ009809 AQ406989 AQ000416 AQ263687
AQ567525 AQ134163 AQ0705263 AQ629490 AQ79378
AQ733071 AQ485474 AQ374766 AQ041660 AQ392400 B90870
AQ368709 AQ392414 AQ416338 AQ518037 AQ122336 AQ534392
AQ353324 B46776 AQ631289 AQ587966 AQ356012 AQ830605 B86437
AQ103375 AQ336818 AQ79135 AQ475699 AQ379051 AQ183111
AQ226770 B42652 AQ371549 AQ456929 AQ213075 AQ486054"
complement(2232..2332)..(2304..2382)
/feature="match: ESTs AW019895 AA810343 AL047793 A1027504
H47919 AA376278 AA335765 AA376189 WA2477 AA053573 AA375658
AL048534 W93437 AA371935 AA490123 AA402704 T55145 AA486656
AA773990 AA283630 H67233 T59499 R50612 AA515631 AA523086
AA302982 AA480492 AA774199 T62882 AA744303 AA015725 T76991
AA558896 AA229904 AA483754 A1816100 A1355803 AA679532
AA745348 R19969 AA290966 AL046262 AA715348 AA234804
AA344085 H51618 F03310 AL120083"
(2252..2345)..(2304..2386)
/feature="match: ESTs AW051819 H72628 A1608668 A1032875
AA181987 AA486559 AA244397 A1669239 A1043105 H49262

```

```

AI023401 AA761008 AA605142 AA829847 AA021034 AA310191
H09123 AA559030 A1871724 T72674 T98886 H81552 AA923712
I16068 AI076357 AT272192 A1368252 H61034 A1051619 H79348
AI473949 AA878105 AI832910 AA483936 AI083610 AA745638
AI310556 AL035988 AA385780 A1168100 AA877973 A1324602
AA878106 AA225230 A1453239 A11859 AA582842 A1953769
AI250409 R36445 AA180344 AA569591 AA631359 A1213714
AA643211 A1978999 AA426021 AA229905 A1168101 AA682635
AA716522 AA019068 A1813484 AL038528 N57913 AA180487
AA252596 AA315566 AL045829 AA100431 AA226153 AA775332
AA809926 AA487071 A1683635 AA469242 C75350 T93052 AA046440
AA642464 AA708753 AA534054 T17218 H61906 A1568836 AA90694
H90784 AA01590 AA720979 H84764 R02522 AA669369 H73442
AA614144 AA732435 AA604995 A1590789 AA659613 AA491954
AA526639 A1113159 AW082490 AW006930 AW006919 AW007621
AW078909 A1992193 AW007989 AL045954 AA643964"
(2252..2341)..(2311..2397)
/feature="match: GSSs AQ020681 AQ422183 AQ395382 AQ543764
B92811 AQ103217 AQ525868 AQ032940 AQ528253 AQ662555
AQ136985 AQ190729 AQ472293 AQ267937 AQ473182 AQ551390
AQ527731 AQ042077 AQ484780 AQ550517 AQ167022 B87081
AQ269024 AQ630192 AQ007744 AQ176782 AQ198723 AQ701150
AQ175828 AQ506370 AQ350441 AQ143462 AQ53970 AQ385987
AQ208962 AQ347476 AQ745990 AQ556682 AQ529563 AQ490365
B68221 AQ530224 B05868 AQ490786 AQ054307 AQ507905 B60385
AQ681971 AQ489261 AQ343157 AQ369527 AQ310151 AQ737169
AQ549438 AQ829593 AQ231224 AQ478614 AQ375510 AQ227632
AQ134004 AQ429116 AQ207026 AQ780082 AQ719813 AQ002146
AQ115286 AQ570005"
2286..2345
/feature="Genefinder prediction"
2388..2402
/feature="ATTT repeat"
complement(2399..2512)
/feature="86% identity: matches 11..124 of consensus"
/rpt_family="AluJo"
2782..2847
/feature="match: EST W91939"
complement(2798..2965)
/feature="Genefinder prediction"
complement(2842..2965)
/feature="GRAIL, score = 79%, comment = excellent shadow"
2967..2979
/feature="TgTC repeat"
3032..3073
/feature="match: GSS AQ535294"
3034..3085
/feature="match: EST A1160319"
complement(3043..3076)
/feature="match: EST AA642169"
3514..3581
/feature="XPOUND prediction, score = 0.528"
3707..3732
/feature="IR1, 84% complementary to IR1' (3750..3775)"
/rpt_type="INVERTED"
3750..3775
/feature="IR1', 84% complementary to IR1 (3707..3732)"
/rpt_type="INVERTED"
3842..3868
/feature="TC repeat"
3842..3867
/feature="homology = 100.0%, counts = 13"
/rpt_family="tg repeat"
/rpt_type="TANDEM"
complement(4101..4141)
/feature="match: EST AA482201"
4104..4180
/feature="GENSCAN prediction, score = 3.89; GRAIL, score =
100%, comment = excellent; MZEF prediction, score = 0.996"
4138..4180
/feature="XPOUND prediction, score = 0.939"
4258..4604
/feature="match: GSS AQ728019"

```

```
misc_feature      4370..4708
/note="match: EST H69816"
/feature="match: GSSs A0542359 A0544443 A0426467 A0126259
/feature="match: GSSs A0542359 A0322830 A0195430 A0337507
A020063 A0546084 A0205053 A0322830 A0195430 A0337507
A021567 A0065988 A0594678 A0036247 A0542440 B34620"
(4786-4813)..(4836-4885)
/feature="match: ESTs A1243353 A1357315 A0325371 A1338348
/feature="match: ESTs A1243353 A1357315 A0325371 A1338348

Query Match      13.9%; Score 19; DB 10; Length 135855;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 AGCTTCCAAAGCTAAACA 98
|||||
Db 16163 AGCTTCCAAAGCTAAACA 16181

RESULT 14
AC005951/c
LOCUS
DEFINITION Homo sapiens chromosome 17, clone hRPK.372_K_20, complete sequence.
ACCESSION AC005951
VERSION AC005951.1 GI:3892081
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155450) PRI 18-NOV-1998
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.372_K_20
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155450)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Collangelo,M., Collins,S.,
Collimore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Hearford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meidrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (07-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 155450)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Collangelo,M., Collins,S.,
Collimore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Hearford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meidrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.
Direct Submission
Submitted (18-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 18, 1998 this sequence version replaced gi:3868745.
All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
Location/Qualifiers
Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.372_K_20"
/clone_lib="RPCI-11 human BAC library"
/map="17"
/chromosome="17"
repeat_region 1968..2273
/rpt_family="AluY"
complement(3334..3431)
/rpt_family="MIR"
repeat_region 3490..3702
/rpt_family="MER58A"
complement(3712..3812)
/rpt_family="L2"
repeat_region 3841..4131
/rpt_family="MLT1D"
complement(4198..4238)
/rpt_family="(TAAA)n"
repeat_region 4239..4522)
/rpt_family="AluJo"
complement(4523..4612)
/rpt_family="L2"
repeat_region 6026..6172
/rpt_family="MIR"
complement(6989..7017)
/rpt_family="(CA)n"
repeat_region 7219..8061)
/rpt_family="L2"
repeat_region 10963..12285
/rpt_family="L1PA4"
repeat_region 12295..12334
/rpt_family="(TAAA)n"
repeat_region 12594..12698
/rpt_family="MLT1J"
complement(13005..13180)
/rpt_family="MIR"
repeat_region 15210..15265)
/rpt_family="L2"
repeat_region 15552..17481)
/rpt_family="L1PA2"
repeat_region 17479..18391
/rpt_family="L1PA2"
complement(19533..19555)
/rpt_family="(CAAA)n"
complement(19556..19838)
/rpt_family="AluSg"
repeat_region 20101..20121
/rpt_family="AT-rich"
repeat_region 20135..20462
/rpt_family="MLT1A1"
repeat_region 21131..22493
/rpt_family="L1MB8"
repeat_region 23468..23817
/rpt_family="L1MB8"
repeat_region 24700..25033
/rpt_family="L1M4"
25053..25172
/feature="Single-stranded terminator coverage."
repeat_region 25183..25677
/rpt_family="L1M4"
repeat_region 25748..25856
/rpt_family="MLT1J"
repeat_region 25958..26130
/rpt_family="MLT1I"
complement(26155..27023)
/rpt_family="L1P"
repeat_region 27025..27424
/rpt_family="L1PA13"
repeat_region 28022..28136
```

```

repeat_region /rpt_family="L1ME2"
28137..28444
repeat_region /rpt_family="AluY"
28445..28692
repeat_region /rpt_family="L1ME2"
28717..28761
repeat_region /rpt_family="AT_rich"
28876..29000
repeat_region /rpt_family="TA)n"
complement(29647..29768)
repeat_region /rpt_family="L2"
30106..31622
repeat_region /rpt_family="L1PA2"
complement(31743..31945)
repeat_region /rpt_family="MLT1b"
complement(32064..32228)
repeat_region /rpt_family="AluJo/FRAM"
complement(32229..32359)
repeat_region /rpt_family="AluSc"
complement(32361..32902)
repeat_region /rpt_family="LTR12"
complement(32960..33008)
repeat_region /rpt_family="Alu"
complement(33159..33195)
repeat_region /rpt_family="(TAA)n"
complement(33196..33476)
repeat_region /rpt_family="AluSc"
34002..34353
repeat_region /rpt_family="MLT1A1"
complement(34355..34772)
repeat_region /rpt_family="MSTA"
complement(34874..34912)
repeat_region /rpt_family="(CAAAA)n"
complement(34913..35185)
repeat_region /rpt_family="AluSg"
35582..35650
repeat_region /rpt_family="AT_rich"
complement(36881..37168)
repeat_region /rpt_family="AluJo"
complement(37211..37558)
repeat_region /rpt_family="L1MB6"
38116..38245
repeat_region /rpt_family="L2"
38406..38446
repeat_region /rpt_family="AT_rich"
38474..40794
repeat_region /rpt_family="L1PA10"
41093..41115
repeat_region /rpt_family="AT_rich"
complement(41286..41333)
repeat_region /rpt_family="purine-rich"
41353..41386
repeat_region /rpt_family="(CA)n"
41781..41882
repeat_region /rpt_family="(TA)n"
42039..42080
repeat_region /rpt_family="AT_rich"
42439..42733
repeat_region /rpt_family="AluSx"
complement(43412..43600)
repeat_region /rpt_family="LTR41"
43822..43857
repeat_region /rpt_family="AT_rich"
43900..44072
repeat_region /rpt_family="L1MD1"
44151..44595
repeat_region /rpt_family="L1MD1"
44772..44816
repeat_region /rpt_family="AT_rich"
45430..45692
repeat_region /rpt_family="L1"
45693..46518
repeat_region /rpt_family="L1MA6"

```

```

repeat_region complement(46617..46812)

Query Match 13.9%; Score 19; DB 40; Length 155450;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGATG 19
Db 52813 CCACCGCACCTGGCTGATG 52795

RESULT 15
AC007092 157176 bp DNA PRI 23-OCT-1999
LOCUS Homo sapiens BAC clone NH0090D01 from 2, complete sequence.
AC007092
AC007092.4 GI:6102673
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 157176)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 157176)
AUTHORS McPherson,C., Strommatt,C. and Davidson,T.
TITLE The sequence of Homo sapiens BAC clone NH0090D01
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 157176)
AUTHORS Waterston,R.H.
TITLE Direct Submisson
JOURNAL Submitted (16-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 157176)
AUTHORS Waterston,R.H.
TITLE Direct Submisson
JOURNAL Submitted (29-JUL-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 157176)
AUTHORS Waterston,R.
TITLE Direct Submisson
JOURNAL Submitted (23-OCT-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 23, 1999 this sequence version replaced gi:5649381.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapienswatson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis

```



MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.5

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is NH0179A20; the clone sequenced to the right is NH0260K08, 200 bp overlap. Actual end of this clone is at base position 15376 of NH0260K08.

## FEATURES

source

Location/Qualifiers

1. 157176

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="NH0090D01"

/clone\_lib="RPCI-11"

1692..2043

/rpt\_family="MaLR"

2738..2780

/rpt\_family="MER1\_type"

2784..2965

/rpt\_family="MER1\_type"

2971..3306

/rpt\_family="MER2\_type"

4916..5060

/rpt\_family="MER1\_type"

5549..5586

/rpt\_family="(TAGA)n"

5615..5801

/rpt\_family="L1"

6812..6880

/rpt\_family="MIR"

6881..7056

/rpt\_family="MaLR"

7057..7343

/rpt\_family="Alu"

7346..7367

/rpt\_family="(TTAA)n"

7388..7596

/rpt\_family="MaLR"

7597..7738

/rpt\_family="MIR"

7769..7803

/rpt\_family="(CTA)n"

8098..8126

/rpt\_family="AT-rich"

8137..8228

/rpt\_family="MIR"

10364..10684

/rpt\_family="Alu"

11309..11456

/rpt\_family="L2"

12538..12558

/rpt\_family="AT-rich"

12564..12620

/rpt\_family="MER1\_type"

12639..12981

/rpt\_family="MER1\_type"

14025..14330

/rpt\_family="Alu"

16136..16267

/rpt\_family="MIR"

16316..16404

/rpt\_family="MaLR"

repeat\_region 16943..17093  
/rpt\_family="MER103"  
repeat\_region 17233..17610  
/rpt\_family="MaLR"  
repeat\_region 17611..19152  
/rpt\_family="MaLR"  
repeat\_region 19153..19528  
/rpt\_family="MaLR"  
repeat\_region 19575..19655  
/rpt\_family="MIR"  
repeat\_region 19867..20055  
/rpt\_family="MaLR"  
repeat\_region 20177..20207  
/rpt\_family="AT-rich"  
repeat\_region 20328..20464  
/rpt\_family="MIR"  
repeat\_region 21588..21816  
/rpt\_family="MIR"  
repeat\_region 22905..22954  
/rpt\_family="(CA)n"  
repeat\_region 24488..24517  
/rpt\_family="(TTTA)n"  
repeat\_region 24527..24732  
/rpt\_family="Alu"  
repeat\_region 25441..25468  
/rpt\_family="AT-rich"  
repeat\_region 26918..27105  
/rpt\_family="MER1\_type"  
repeat\_region 28798..28976  
/rpt\_family="MIR"  
repeat\_region 28988..29865  
/rpt\_family="L1"  
repeat\_region 29916..30050  
/rpt\_family="MER1\_type"  
repeat\_region 31276..31397  
/rpt\_family="MER1\_type"  
repeat\_region 31548..31648  
/rpt\_family="MER1\_type"  
repeat\_region 32045..32282  
/rpt\_family="MER1\_type"  
repeat\_region 32289..32398  
/rpt\_family="MIR"  
repeat\_region 32505..32590  
/rpt\_family="(TC)n"  
repeat\_region 33437..33467  
/rpt\_family="(GGGTG)n"  
repeat\_region 36487..36538  
/rpt\_family="(CA)n"  
repeat\_region 38116..38143  
/rpt\_family="(TC)n"  
repeat\_region 38143..38179  
/rpt\_family="(TG)n"  
repeat\_region 38629..38988  
/rpt\_family="MaLR"  
repeat\_region 39148..39426  
/rpt\_family="Alu"  
repeat\_region 39478..39699  
/rpt\_family="MER1\_type"  
repeat\_region 39722..39891  
/rpt\_family="MaLR"  
repeat\_region 40336..40455  
/rpt\_family="MER94"  
repeat\_region 40564..40725  
/rpt\_family="MaLR"  
repeat\_region 40836..40967  
/rpt\_family="MaLR"  
repeat\_region 41585..41624

Query Match

Best Local Similarity 13.9%; Score 19; DB 40; Length 157176;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGATG 19

```

|||||
Db 82915 CCACCGCAGCTGGCTGATG 82933

RESULT 16
AC012536/c
LOCUS
DEFINITION
AC012536
VERSION
KEYWORDS
SOURCE
ORGANISM

AC012536 160798 bp DNA HTG 30-NOV-1999
Homo sapiens clone RP11-114J22, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
AC012536 GI:6479135
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160798)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-114J22
Unpublished
2 (bases 1 to 160798)
Balgwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhealter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (29-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 30, 1999 this sequence version replaced gi:6139161.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L1935
Center clone name: 114_J_22
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 138554 bases at least Q40
Consensus quality: 131363 bases at least Q30
Consensus quality: 157431 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 160798; sum-of-ctgts
Quality coverage: 4.9 in Q20 bases; agarose-fp
Quality coverage: 5.0 in Q20 bases; sum-of-ctgts
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1039: contig of 1039 bp in length
1040 2334: gap of unknown length
2334: contig of 1295 bp in length

```

```

* 2335 4510: contig of 2176 bp in length
* 4511 18352: contig of 13842 bp in length
* 18353 29917: contig of 11565 bp in length
* 29918 50970: contig of 21053 bp in length
* 50971 75859: contig of 24889 bp in length
* 75860 112001: contig of 36142 bp in length
* 112002 160798: contig of 48797 bp in length.
FEATURES
Location/Qualifiers
1..160798
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RPCI-11 Human Male BAC"
BASE COUNT 51460 a 28275 c 27976 g 53079 t 8 others
ORIGIN
Query Match 13.9%; Score 19; DB 44; Length 160798;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCACCGCAGCTGGCTGATG 19
|||||
Db 148882 CCACCGCAGCTGGCTGATG 148864
RESULT 17
LOCUS
DEFINITION
HSL179N16 172048 bp DNA PRI 23-NOV-1999
Homo sapiens DNA sequence from PAC 179N16 on chromosome
6p21.1-21.33. Contains the SAPK4 (MAPK p38delta) gene, and the
alternatively spliced SAPK2 gene coding for CSaids binding protein
CSBP2 and a MAPK p38beta LIKE protein. Contains ESTs, STSs and two
predicted CpG islands, complete sequence.
Z95152
Z95152.1 GI:3036773
HTG: CpG island; CSaids binding protein; CSBP2; MAPK; Mitogen
activated protein kinase; MXI2; P38; SAPK4; Stress activated
protein kinase.
human.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172048)
Palmer,S.
Direct Submission
Submitted (02-APR-1998) Chromosome 6 Project Group
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquyres@sanger.ac.uk
On Apr 8, 1998 this sequence version replaced gi:2505919.
IMPORTANT: This sequence is the entire insert of clone 179N16.
During sequence assembly data are compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed in collaboration by the Sanger
Centre chromosome 6 mapping group and Ioannis Ragousis. Further
information can be found at http://www.sanger.ac.uk/HGP/Chr6/
This sequence has been finished according to sequence map criteria as
follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the

```



27782. .27861.29276. .29347.29456. .29570.29695. .29742,  
30178. .30207.30365. .30473.33547. .33605.34826. .34955,  
35525. .35643))  
/gene="sAPK4"  
/note="match: proteins O14739 O15124"  
/codon\_start=1  
/evidence=not\_experimental  
/product="dj179N16.1 (Stress Activated Protein Kinase 4  
(Mitogen Activated Protein Kinase p38delta))"  
/protein\_id="CAB08438.1"  
/db\_xref="GI:3036774"  
/db\_xref="SPTREMBL:O15264"  
/translation="MSLRKKGFYKQDVNKTAWELPKTYVSPTHVSGAGSVCSAID  
KRSGEVAIKLSPFQSEIFAKRAYRELLIKHMENHVGILDFVTASSLRNFYD  
FYLVPMTDLQKIMGEFSEEEKIQLYVQMLKGLKYHSAGVVHRLKPGNLAVNE  
DCELKILDFGLARHAAEMTGYVVRWYRAFEVILSWMHYNOTVDINSVGCMAEMLT

Query Match 13.9%; Score 19; DB 10; Length 172048;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGATG 19  
|||||  
Db 41760 CCACCGCACCTGGCTGATG 41742

RESULT 18  
AC007158  
LOCUS AC007158 204027 bp DNA HTG 26-OCT-1999  
DEFINITION Homo sapiens clone hRPK.90\_A\_1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2  
ordered pieces.  
ACCESSION AC007158  
VERSION AC007158.5 GI:5757580  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 204027)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens, clone hRPK.90\_A\_1  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 204027)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Hearford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Rilev,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Rhmann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.  
DIRECT SUBMISSION  
TITLE Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome  
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Aug 22, 1999 this sequence version replaced gi:4726136.  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and

\* the accession number will be preserved  
\* 1 160375; contig of 160375 bp in length  
\* gap of unknown length  
\* 160376 204027; contig of 43652 bp in length.

FEATURES  
source  
Location/Qualifiers  
1..204027  
/Organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="hRPK.90\_A\_1"  
/clone\_lib="rPCI-11 Human Male BAC"  
BASE COUNT 63958 a 40107 c 39047 g 60875 t  
ORIGIN  
Unreported.

Query Match 13.9%; Score 19; DB 42; Length 204027;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACGGCACCTGGCTGATG 19  
|||||  
Db 111376 CCACGGCACCTGGCTGATG 111394

RESULT 19  
HUMVIPHMI  
LOCUS HUMVIPHMI 270 bp DNA PRI 14-JAN-1995  
DEFINITION Human vasoactive intestinal peptide and histidine-methionine amide  
peptide hormone (VIP/PHM-27) gene, exon 1.  
ACCESSION M14618  
VERSION M14618.1 GI:340266  
KEYWORDS histidine-methionine amide; hormone; peptide hormone; vasoactive  
intestinal peptide.  
SEGMENT 1 of 6  
SOURCE Human DNA (libraries of Lawn and EMBL), clones lambda-VIP-D and  
lambda-VIP-4.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 270)  
AUTHORS Linder,S., Barkhem,T., Norberg,A., Persson,H., Schalling,M.,  
Hokfelt,T. and Magnusson,G.  
TITLE Structure and expression of the gene encoding the vasoactive  
intestinal peptide precursor  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84 (2), 605-609 (1987)  
MEDLINE 87092456  
COMMENT Draft entry and clean copy of sequence [1] kindly provided by  
S.Linder, 02-MAR-1987.

FEATURES  
source  
Location/Qualifiers  
1..270  
/Organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="6q24-q27"  
prim\_transcript 100..>270  
/gene="vip"  
/note="VIP/PHM27 mRNA; G00-120-490"  
intron 264..>270  
/gene="vip"  
/note="VIP/PHM27 mRNA intron A"  
BASE COUNT 76 a 67 c 69 g 58 t  
ORIGIN  
Unreported.

Query Match 13.1%; Score 18; DB 9; Length 270;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 TGACTTCCTTCAGAGGAC 47  
|||||  
Db 22 TGACTTCCTTCAGAGGAC 39

RESULT 20  
G24450

LOCUS G24450 474 bp DNA STS 31-MAY-1996  
 DEFINITION human STS WI-13939, sequence tagged site.  
 ACCESSION G24450  
 VERSION G24450.1 GI:1344776  
 KEYWORDS STS; STS sequence; primer; sequence tagged site.  
 SOURCE human STS derived from sequences in dbEST and the Unigene collection.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 474)  
 AUTHORS Hudson, T.  
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically  
 Mapped STS  
 JOURNAL Unpublished (1995)  
 COMMENT

Contact: Thomas Hudson  
 Whitehead Institute/MIT Center for Genome Research  
 Whitehead Institute for Biomedical Research  
 9 Cambridge Center, Cambridge MA 02142 USA  
 Tel: 617 252 1900  
 Fax: 617 252 1902  
 Email: thudson@genome.wi.mit.edu

Primer A: CTCCTCTTTTCATTACCCAC  
 Primer B: GTGACAGAGTCTTCCGAGG  
 STS size: 150  
 PCR Profile:

presoak:  
 Denaturation: 56 degrees C  
 Annealing: 56 degrees C  
 Polymerization: 35  
 PCR Cycles: 35  
 Thermal Cycler:  
 Protocol:  
 Template: 10 ng  
 Primer: each 5 pm  
 dNTPs: each 4 mM  
 Taq Polymerase: 0.025 units/ul  
 Total Vol: 20 ul

Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 9.3

Derived from dbEST (genbank accession T17065).

FEATURES  
 source  
 Location/Qualifiers  
 1..474  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="441.4 cR from top of Chr17 linkage group"  
 STS  
 primer\_bind  
 1..150  
 primer\_bind  
 1..20  
 complement(131..150)  
 BASE COUNT 127 a 117 c 101 g 123 t 6 others  
 ORIGIN  
 Query Match 13.1%; Score 18; DB 13; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 52 AAAGACACTAAGTGGAA 69  
 |||||  
 Db 258 AAAGACACTAAGTGGAA 275  
 |||||  
 Query Match 13.1%; Score 18; DB 13; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 6.6;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 52 AAAGACACTAAGTGGAA 69  
 |||||  
 Db 258 AAAGACACTAAGTGGAA 275  
 |||||

RESULT 21  
 HSU66615/c mRNA PRI 18-SEP-1996  
 LOCUS HSU66615 5190 bp  
 DEFINITION Human SWI/SNF complex 155 Kda subunit (BAF155) mRNA, complete cds.

U66615  
 U66615.1 GI:1549238  
 human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 5190)  
 AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R. and Crabtree, G.R.  
 TITLE Diversity and specialization of mammalian SWI/SNF complexes  
 JOURNAL Genes Dev. 10 (17), 2117-2130 (1996)  
 MEDLINE 96397413  
 REFERENCE 2 (bases 1 to 5190)  
 AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R. and Crabtree, G.R.  
 TITLE Purification and biochemical heterogeneity of the mammalian SWI-SNF complex  
 JOURNAL EMBO J. 15 (1996) In press  
 REFERENCE 3 (bases 1 to 5190)  
 AUTHORS Wang, W., Xue, Y., Zhou, S. and Crabtree, G.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA  
 FEATURES  
 Location/Qualifiers  
 1..5190  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="Jurkat T cells"  
 55..3369  
 /gene="BAF155"  
 55..3369  
 /gene="BAF155"  
 /note="a core subunit presents in all human SWI/SNF complexes purified so far; similar to subunit BAF170 and yeast Swi3 protein; contains a region similar to DNA binding domain of myb and a predicted leucine zipper; the C-terminus of the protein is highly proline-rich and somewhat glutamine-rich"  
 /codon\_start=1  
 /product="SWI/SNF complex 155 Kda subunit"  
 /protein\_id="AAC50893.1"  
 /db\_xref="GI:1549239"  
 /translation="MAAAGGGGPGTAVGATGFGDSAAAGLAVYRRKDGPGATKFEW  
 SPETVQLDSRVVWLGKHYKVHADPTNKTLAGLVOLLOFOEDAFGKHVTPAFT  
 KLPAKFMDFKAGALCHILGAAYKYKNEQWRERDLQPSRMDRVNMFMTETLV  
 ONCLAFPNLYLPIDDLKANKLDIIRHQCTTDEKSKASHLYPYSSQDDSEW  
 LRPVMRKEQVLVHNGFYDSDYTWHSNDVDRAIEDPTPEKPKVHKVHWTDDIF  
 NEWNEEDYEDENRKPVSFRQISTKNEEPVRSRRDRKASARKRKHSPSPPPP  
 TPTESRKKSGKQASLYGKRKQKEDEDLTKDMEDPTVPNTEEVYLPKNVNLK  
 KDSNTPVKGGTVADLDEODEETVTAGGDEDEPAKGDQSRSDVLGDNVBTQTHII  
 IPSYASWFYDNCIHTERRALPEFFGNKSKTPELYLAVRNFMDISYRLNPOEYLT  
 TACRNLITGDVCAVMRVHAGGEOMLVNYQVDPESRPMAGPPTPHFNVLADTPLAC  
 ASDLRSPQVPAQOMLNFPEKNEKRPVDLQNGFLRTDIYSKTLAKSKASAGRWTE  
 QETLLLEALMYKDDNDKVSERHVSQTODECILHFLRUPIDPFIENSDASUGPLAY  
 OPVPFSQGNPVMSTVAFASVYDPRVSAAKAALEEFSSRVREVEYPLDELVAHKV  
 QEAARASGVDPPTYGLESCVAGTGPDEPEKLEGAEEKMEADPDQOQPEAKNKV  
 ETDEGKAODGENEKSEKQSEVSDTKESEKTEENKELSTCKRESDTGKKV  
 EHEISEGNVATAAALASATKAKHLAAVEERKIKSLVALLVETOMKLEIKLRHFE  
 GLETIMDRXEALEQOQOLLTERONFHEQLYAEILRARQOQNEQOHCNQOAHQH  
 SGGFGLAPLGRAGHPGMHQPPPIPLMHQMPHPPPQGOIPGPGSMGPHQHPG  
 RMTPTVAANTHPSSGPTPPGMPMPGNILGRVPLTAPNGMYPPPPQOQPPPPPAD  
 GVPPPPAGGPPASAAP"  
 BASE COUNT 1461 a 1168 c 1297 g 1264 t  
 ORIGIN  
 Query Match 13.1%; Score 18; DB 11; Length 5190;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 13 GCTGATGCTTTCTATCT 30  
 |||||

```
Db 1013 GCTGATGCTTTTCTATCT 996

RESULT 22
AC008821/c
LOCUS AC008821 14837 bp DNA HTG 03-AUG-1999
DEFINITION Homo sapiens chromosome 5 clone CITB-HL_2129G21, *** SEQUENCING IN
PROGRESS ***, 18 unordered pieces.
ACCESSION AC008821
VERSION AC008821.1 GI:5686250
KEYWORDS HTG; HIGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 14837)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14837)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 649: contig of 649 bp in length
* 650 1303: contig of 654 bp in length
* 1304 1962: contig of 659 bp in length
* 1963 2611: contig of 649 bp in length
* 2612 2733: contig of 122 bp in length
* 2734 3378: contig of 645 bp in length
* 3379 4036: contig of 658 bp in length
* 4037 4670: contig of 634 bp in length
* 4671 5597: contig of 927 bp in length
* 5598 6302: contig of 705 bp in length
* 6303 7324: contig of 1022 bp in length
* 7325 7966: contig of 642 bp in length
* 7967 8984: contig of 1018 bp in length
* 8985 9882: contig of 898 bp in length
* 9883 10526: contig of 644 bp in length
* 10527 11913: contig of 1387 bp in length
* 11914 13303: contig of 1390 bp in length
* 13304 14837: contig of 1334 bp in length.
Location/Qualifiers
1..14837
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"

FEATURES
source
Location/Qualifiers

BASE COUNT 4433 a 3077 c 3067 g 4254 t 6 others
ORIGIN

Query Match 13.18; Score 18; DB 41; Length 14837;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACGGCACCTGGCTGAT 18
|||||
Db 12740 CCACGGCACCTGGCTGAT 12723

RESULT 23
AC002120/c
LOCUS AC002120 23536 bp DNA PRI 31-JUL-1998
DEFINITION Homo sapiens chromosome 9q34, clone 182E8, complete sequence.
ACCESSION AC002120
VERSION AC002120.1 GI:3366591
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 23536)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 9q34, clone 182E8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 23536)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fritpp,W.J.,
Gage,D., Geraghty,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., Mcdermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (14-MAY-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 23536)
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C.,
Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B.,
Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye.W.J., Zhao,J. and
Zody,M.
Direct Submission
Submitted (31-JUL-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 31, 1998 this sequence version replaced gi:3184311.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only 23.5 kilobases of this clone are being submitted. The
remainder overlaps either accession number AC000393 (WICGR project
L174) or accession number AC000394 (WICGR project L177).
```

```
Source
1. .23536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="182E8"
/clone_lib="Dr. David Kwiatkowski"
/map="9c34"
/chromosome="9"
complement(215. .321)
/rpt_family="MIR"
1680. .1774
/rpt_family="L2"
1892. .2023
/rpt_family="AluJo"
2024. .2341
/rpt_family="AluSg"
2414. .2448
/rpt_family="GAAAn"
2754. .2888
/rpt_family="AluSq/x"
2890. .3185
/rpt_family="AluY"
3188. .3325
/rpt_family="AluSg/x"
3493. .3520
/rpt_family="(CA)n"
4310. .4406
/rpt_family="(GA)n"
complement(5200. .5496)
/rpt_family="AluX"
complement(5828. .5928)
/rpt_family="L2"
5978. .6111
/rpt_family="AluSq/x"
6112. .6413
/rpt_family="AluSp"
complement(6782. .6812)
/rpt_family="(GAA)n"
complement(6824. .7114)
/rpt_family="AluJb"
complement(7117. .7337)
/rpt_family="AluJo"
complement(8457. .8514)
/rpt_family="L2"
10625. .11063
/rpt_family="L1MB3"
11366. .11503
/rpt_family="L2"
11681. .11862
/rpt_family="(TGGAn)"
11878. .11996
/rpt_family="(TGGAn)"
11998. .12100
/rpt_family="(TGGAn)"
complement(12478. .12512)
/rpt_family="MIR"
complement(12789. .12913)
/rpt_family="(GGA)n"
13842. .14006
/rpt_family="L1MC2"
complement(14203. .14225)
/rpt_family="(TA)n"
14992. .15160
/rpt_family="L1ME2"
15187. .15494
/rpt_family="AluJb"
15509. .15936
/rpt_family="L1M4"
15939. .16234
/rpt_family="AluX"
16236. .16795
/rpt_family="L1M4"
16796. .17086
/rpt_family="AluSg"

repeat_region 17088. .17250
/rpt_family="AluJb"
repeat_region 17277. .17581
/rpt_family="AluX"
repeat_region 17723. .18068
/rpt_family="L1ME3A"
repeat_region 18073. .18370
/rpt_family="AluX"
unsure complement(18154. .18174)
/rpt_family="Single-stranded coverage."
repeat_region 18372. .18638
/rpt_family="AluJb"
repeat_region 18643. .19029
/rpt_family="L1ME3A"
repeat_region complement(19050. .19522)
/rpt_family="MLTIG"
unsure 19077. .19167
/rpt_family="Single-stranded coverage."
unsure complement(19559. .19580)
/rpt_family="Single-stranded coverage."
repeat_region 19632. .19779
/rpt_family="MLTIA2"
repeat_region 19781. .19878
/rpt_family="AluSg/x"
repeat_region 19888. .20108
/rpt_family="MLTIA2"
repeat_region 20109. .20435
/rpt_family="L1ME3A"
repeat_region 20436. .20736
/rpt_family="AluX"
repeat_region 20738. .20795
/rpt_family="L1ME3"
repeat_region 20805. .21205
/rpt_family="L1M4"
unsure complement(21059. .21207)
/rpt_family="Single-stranded terminator coverage."
repeat_region complement(21284. .21555)
/rpt_family="AluSg"
repeat_region 21579. .22388
/rpt_family="TIGGER1"
unsure 21582
/rpt_family="Uncertain number of A's; either 24, 25, or 26 A's."
unsure complement(22159. .22762)
/rpt_family="Single-stranded terminator coverage."
repeat_region 22389. .22685
/rpt_family="AluSp"
repeat_region 22686. .22994
/rpt_family="TIGGER1"
repeat_region 22996. .23288
/rpt_family="AluJb"
repeat_region 23291. .23458
/rpt_family="TIGGER1"
repeat_region complement(23460. .23493)
/rpt_family="L1MB3"
repeat_region complement(23500. .23535)
/rpt_family="Alu"
BASE COUNT 6457 a 5475 c 6200 g 5404 t
ORIGIN

Query Match 13.1%; Score 18; DB 11; Length 23536;
Best Local Similarity 100.0%; Pred.No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGACCTGGCTGAT 18
|||||
Db 23009 CCACCGACCTGGCTGAT 22992

RESULT 24
U73643/c 33602 bp DNA PRI 24-JUL-1997
LOCUS U73643 Human Chromosome 11 Cosmid cSHL34e5, complete sequence.
DEFINITION
```

```

ACCESSION      U73643
VERSION        U73643.1  GI:2276400
KEYWORDS       HTG.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 33602)
AUTHORS        Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
                Davies,C.J., Davis,C., English,C., Fondon,T.,
                Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
                Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U.,
                Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
                Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
                Ward,T. and Wilson,R.
TITLE          Template
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 33602)
AUTHORS        Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
                Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
                Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hanher,L., Harris,J.,
                Hinson,S., McFarland,J., Megarity,C., Narayanaswamy,U., Newton,J.,
                O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
                Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
                Ward,T. and Wilson,R.
TITLE          Direct Submission
JOURNAL        Submitted (07-OCT-1996) McDermott Center for Human Growth and
                Development, University of Texas Southwestern Medical Center at
                Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE      3 (bases 1 to 33602)
AUTHORS        Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Burbee,D.,
                Davie,J., Davies,C.J., Davis,C., English,C., Fondon,T.,
                Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O.,
                Hahner,L., Harris,J., Hinson,S., Megarity,C., Narayanaswamy,U.,
                Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S.,
                Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
                Ward,T. and Wilson,R.
TITLE          Direct Submission
JOURNAL        Submitted (24-JUL-1997) Genome Science and Technology Center,
                University of Texas Southwestern Medical Center at Dallas, 5323
                Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT        On Jul 24, 1997 this sequence version replaced gi:1737198.
FEATURES       Location/Qualifiers
                source          1..33602
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /clone="CSR134es"
                                /chromosome="11"
                                complement(448..727)
                                /rpt_family="Alu"
                                1040..1365
                                /rpt_family="Alu"
                                complement(2027..2078)
                                /rpt_family="Alu"
                                complement(2289..2841)
                                /rpt_family="Alu"
                                complement(2890..3084)
                                /rpt_family="Alu"
                                complement(3654..3938)
                                /rpt_family="Alu"
                                complement(4104..4350)
                                /rpt_family="Alu"
                                5820..6110
                                /rpt_family="Alu"
                                6336..6591
                                /rpt_family="Alu"
                                6852..7070
                                /rpt_family="MIR"
                                complement(7083..7358)
                                /rpt_family="Alu"
                                complement(7621..7793)
                                /rpt_family="MIR"
                                8519..8779

repeat_region      /rpt_family="Alu"
                    9325..9613
repeat_region      /rpt_family="Alu"
                    10632..10680
repeat_region      /rpt_family="MIR"
                    complement(10811..11084)
repeat_region      /rpt_family="Alu"
                    11827..12092
repeat_region      /rpt_family="Alu"
                    complement(12178..12249)
repeat_region      /rpt_family="MIR"
                    complement(13207..13495)
repeat_region      /rpt_family="Alu"
                    complement(13508..14101)
repeat_region      /rpt_family="Alu"
                    14590..14772
repeat_region      /rpt_family="MIR"
                    16109..16401
repeat_region      /rpt_family="Alu"
                    complement(18386..18965)
repeat_region      /rpt_family="Alu"
                    complement(19286..19873)
repeat_region      /rpt_family="LTR9"
                    19986..20361
repeat_region      /rpt_family="Alu"
                    complement(21464..21765)
repeat_region      /rpt_family="Alu"
                    complement(22974..23553)
repeat_region      /rpt_family="Alu"
                    complement(24505..25021)
repeat_region      /rpt_family="MIR"
                    complement(25246..25504)
repeat_region      /rpt_family="Alu"
                    28658..28950
repeat_region      /rpt_family="Alu"
                    complement(29341..29831)
repeat_region      /rpt_family="L1"
                    30332..30627
repeat_region      /rpt_family="Alu"
                    complement(31736..32027)
repeat_region      /rpt_family="Alu"
                    complement(32818..32919)
repeat_region      /rpt_family="Alu"

BASE COUNT      8284 a 8121 c 8156 g 9041 t
ORIGIN
Query Match      13.1%; Score 18; DB 10; Length 33602;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CCACCGCACCTGGCTGAT 18
        |||||
DB      28664 CCACCGCACCTGGCTGAT 28647

RESULT 25
AC014889/c
LOCUS      AC014889.1  GI:6436446
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION  AC014889
VERSION    AC014889.1  GI:6436446
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 35967)
AUTHORS    Adams,M. and Venter,J.C.
TITLE      Direct Submission
JOURNAL    Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,

```





```
repeat_region 13651..13919
/note="AluY repeat: matches 60. .301 of consensus"
repeat_region 14110..14432
/note="AluY repeat: matches 13. .302 of consensus"
repeat_region 14722..15024
/note="AluX repeat: matches 1. .302 of consensus"
repeat_region 15222..15326
/note="AluY repeat: matches 196. .300 of consensus"
repeat_region 15714..16016
/note="AluY repeat: matches 1. .290 of consensus"
repeat_region 16424..16531
/note="FLAN_A repeat: matches 117. .6 of consensus"
repeat_region 18974..19272
/note="AluS repeat: matches 303. .2 of consensus"
repeat_region 19497..19756
/note="AluY repeat: matches 39. .298 of consensus"
repeat_region 20312..20478
/note="MIR repeat: matches 260. .69 of consensus"
misc_feature 21146..23225
/note="Putative CpG island"
repeat_region 22691..23782
/note="SVA repeat: matches 1101. .1 of consensus"
repeat_region 23783..23828
/note="23 copies of 2 mer 80 % conserved"
repeat_region 24117..24481
/note="L1PA5 repeat: matches 517. .890 of consensus"
repeat_region 25101..25284
/note="MER20 repeat: matches 155. .1 of consensus"
repeat_region 26018..26081
/note="32 copies of 2 mer 83 % conserved"
repeat_region 26083..26258
/note="AluY repeat: matches 195. .5 of consensus"
repeat_region 26463..26760
/note="AluY repeat: matches 3. .300 of consensus"
repeat_region 26761..27058
/note="AluS repeat: matches 2. .302 of consensus"
repeat_region 27676..27988
/note="AluX repeat: matches 1. .298 of consensus"
repeat_region 28413..28685
/note="AluX repeat: matches 298. .2 of consensus"
repeat_region 28718..28848
/note="L1 repeat: matches 3151. .3280 of consensus"
repeat_region 29553..29796
/note="MLT2_internal repeat: matches 5375. .5124 of consensus"
repeat_region 30109..30302
/note="MLT2_internal repeat: matches 4754. .4572 of consensus"
repeat_region 30324..30621
/note="AluX repeat: matches 1. .299 of consensus"
repeat_region 30655..30952
/note="AluS repeat: matches 1. .296 of consensus"
repeat_region 30955..31101
/note="MER4B repeat: matches 11. .148 of consensus"
repeat_region 31637..31915
/note="AluX repeat: matches 1. .287 of consensus"
repeat_region 31977..33195
/note="MLT2_internal repeat: matches 4572. .3316 of consensus"
repeat_region 33199..33502
/note="AluJ repeat: matches 1. .302 of consensus"
repeat_region 33616..33917
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 33984..34529
/note="MLT2_internal repeat: matches 3168. .2597 of consensus"
repeat_region 34644..34800
/note="MLT2_internal repeat: matches 2419. .2264 of consensus"
```

```
repeat_region 35043..35343
/note="AluY repeat: matches 1. .301 of consensus"
misc_feature 35348..35643
/note="match:STS L42688"
repeat_region 35350..35569
/note="MLT2_internal repeat: matches 1972. .1754 of consensus"
Query Match 13.1%; Score 18; DB 11; Length 37027;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 4309 CCACCGCACCTGGCTGAT 4292
|||||

RESULT 27
AC000394/c
LOCUS AC000394 44294 bp DNA PRI 29-MAY-1997
DEFINITION Genomic sequence from Human 9q34, complete sequence.
AC000394
AC000394
AC000394.1 GI:2133911
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 44294)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H. and Lander,E.S.
AUTHORS
Fasman,K.H. and Lander,E.S.
TITLE
Genomic sequence from Human 9q34
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 44294)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (08-APR-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 44294)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frapp,W.J.,
Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,
Marquis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stillwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
Direct Submission
Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 29, 1997 this sequence version replaced gi:1929461.
The Staden databases, finishing information, and all
chromatographic files used in the assembly of this clone are
available from our anonymous ftp site.
All repeats were identified using RepeatMasker: Smit, A.F.A. 6
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
Location/Qualifiers
source
1. .44294
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="81F1"
/clone_lib="Dr. David Kwiatkowski"
/chromosome="9q34"
repeat_region 1. .244
```

```

repeat_region /rpt_family="TIGGER1"
246. .538
repeat_region /rpt_family="AluJb"
541. .708
repeat_region /rpt_family="TIGGER1"
complement(693. .743)
repeat_region /rpt_family="LMB3"
complement(750. .1045)
repeat_region /rpt_family="AluSx"
complement(1054. .1183)
repeat_region /rpt_family="AluJo"
1211. .1502
repeat_region /rpt_family="AluSx"
3258. .3300
repeat_region /rpt_family="(TAA)n"
complement(3306. .3361)
repeat_region /rpt_family="LMB8"
3713. .4016
repeat_region /rpt_family="AluSx"
4017. .4062
repeat_region /rpt_family="POLY_A"
complement(4233. .4254)
repeat_region /rpt_family="AT_rich"
complement(4832. .5135)
repeat_region /rpt_family="AluSx"
5744. .5865
repeat_region /rpt_family="MIR2"
complement(5892. .6211)
repeat_region /rpt_family="MLT1A1"
6460. .6759
repeat_region /rpt_family="AluSg"
complement(6826. .6947)
repeat_region /rpt_family="MLT1B"
7299. .8285
repeat_region /rpt_family="LTR13"
complement(8500. .8858)
repeat_region /rpt_family="MSTD"
complement(8968. .9269)
repeat_region /rpt_family="AluSx"
complement(9566. .9705)
repeat_region /rpt_family="MIR2"
10801. .11099
repeat_region /rpt_family="AluSc"
11111. .11319
repeat_region /rpt_family="AluJo"
complement(11377. .11634)
repeat_region /rpt_family="AluJo"
11817. .12107
repeat_region /rpt_family="AluSx"
12293. .12334
repeat_region /rpt_family="MIR2"
13306. .13462
repeat_region /rpt_family="LTR10"
complement(13467. .13688)
repeat_region /rpt_family="AluSg"
13689. .13988
repeat_region /rpt_family="LTR10"
complement(14257. .14347)
repeat_region /rpt_family="AluSg"
14349. .14514
repeat_region /rpt_family="LTR10"
20090. .20214
repeat_region /rpt_family="LTR10"
complement(20571. .20875)
repeat_region /rpt_family="AluSx"
complement(20980. .21025)
repeat_region /rpt_family="MIR"
complement(21084. .21201)
repeat_region /rpt_family="MIR"
21438. .21513
repeat_region /rpt_family="MIR2"
complement(21522. .21805)
repeat_region /rpt_family="AluJo"

```

```

repeat_region complement(21912. .21978)
/rpt_family="MIR"
repeat_region complement(22248. .22295)
/rpt_family="MIR2"
22898. .23262
repeat_region /rpt_family="LTR13"
23261. .23894
repeat_region /rpt_family="LTR13"
24976. .25119
repeat_region /rpt_family="MIR"
25172. .25264
repeat_region /rpt_family="(GAA)n"
25425. .25480
repeat_region /rpt_family="MER5A"
complement(25534. .25577)
repeat_region /rpt_family="MADEL"
26120. .26286
repeat_region /rpt_family="AluSg"
27580. .27687
repeat_region /rpt_family="MIR2"
complement(27754. .28057)
repeat_region /rpt_family="AluSx"
28165. .28468
repeat_region /rpt_family="AluJo"
complement(29642. .29942)
repeat_region /rpt_family="AluSg"
29964. .30260
repeat_region /rpt_family="AluJb"
30337. .30637
repeat_region /rpt_family="AluY"
30638. .30803
repeat_region /rpt_family="AluSp"
complement(31900. .32128)
repeat_region /rpt_family="LMC1"
32269. .32425
repeat_region /rpt_family="MIR"
complement(34398. .34533)
repeat_region /rpt_family="(GAA)n"
complement(34952. .35259)
repeat_region /rpt_family="AluSx"
complement(35342. .35417)
repeat_region /rpt_family="MIR"
35424. .35718
repeat_region /rpt_family="AluJb"
complement(35731. .35865)
repeat_region /rpt_family="MIR"
39994. .40040
repeat_region /rpt_family="(GAA)n"
complement(40695. .40738)
repeat_region /rpt_family="(TGAA)n"
complement(41319. .41601)
repeat_region /rpt_family="AluJb"
complement(41661. .41736)
repeat_region /rpt_family="MER42C"
complement(41929. .42224)
repeat_region /rpt_family="MER42C"
42271. .42568
repeat_region /rpt_family="AluJb"
42574. .42745
repeat_region /rpt_family="LMC1"
complement(42768. .42853)
repeat_region /rpt_family="(GAA)n"
BASE COUNT 10447 a 11319 c 11448 g 11080 t
ORIGIN

```

Query Match 13.1%; Score 18; DB 10; Length 44294;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGCACCTGGCTGAT 18  
 |||||  
 Db 259 CCACCGCACCTGGCTGAT 242

```

RESULT 28
AC012318/c
LOCUS
DEFINITION
Homo sapiens chromosome 19 clone LLNL-F_143B12, 31-OCT-1999
PROGRESS ***, 23 unordered pieces.
AC012318
AC012318.2 GI:6165034
HTG: HTGS_PHASE1.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 48925)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 19
Unpublished
2 (bases 1 to 48925)
DOE Joint Genome Institute.
Direct Submission
Submitted (23-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:6102621.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
* 1 700: contig of 700 bp in length
* gap of unknown length
*
* 701 3126: contig of 2426 bp in length
* gap of unknown length
*
* 3127 3577: contig of 451 bp in length
* gap of unknown length
*
* 3578 4292: contig of 715 bp in length
* gap of unknown length
*
* 4293 5134: contig of 842 bp in length
* gap of unknown length
*
* 5135 5766: contig of 632 bp in length
* gap of unknown length
*
* 5767 6491: contig of 725 bp in length
* gap of unknown length
*
* 6492 7211: contig of 720 bp in length
* gap of unknown length
*
* 7212 8169: contig of 958 bp in length
* gap of unknown length
*
* 8170 8934: contig of 765 bp in length
* gap of unknown length
*
* 8935 9750: contig of 816 bp in length
* gap of unknown length
*
* 9751 11887: contig of 2137 bp in length
* gap of unknown length
*
* 11888 13847: contig of 1960 bp in length
* gap of unknown length
*
* 13848 15721: contig of 1874 bp in length
* gap of unknown length
*
* 15722 17899: contig of 2178 bp in length
* gap of unknown length
*
* 17900 19846: contig of 1947 bp in length
* gap of unknown length
*
* 19847 21938: contig of 2092 bp in length
* gap of unknown length
*
* 21939 24668: contig of 2730 bp in length
* gap of unknown length
*
* 24669 28196: contig of 3528 bp in length
* gap of unknown length
*
* 28197 30957: contig of 2761 bp in length
* gap of unknown length

```

```

* 30958 34868: contig of 3911 bp in length
* gap of unknown length
* 34869 40884: contig of 6016 bp in length
* gap of unknown length
* 40885 48925: contig of 8041 bp in length.
FEATURES
Location/Qualifiers
source
1..48925
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LLNL-F_143B12"
BASE COUNT 12473 a 12710 c 12124 g 11601 t 17 others
ORIGIN
Query Match 13.1%; Score 18; DB 42; Length 48925;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCACCGCACCTGGCTGAT 18
|||||
Db 45272 CCACCGCACCTGGCTGAT 45255
|||||
RESULT 29
AC011108/c
LOCUS
DEFINITION
Homo sapiens clone 115_K_14, LOW-PASS SEQUENCE SAMPLING.
AC011108
AC011108.1 GI:6006220
HTG: HTGS_PHASE0.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 55707)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone 115_K_14
Unpublished
2 (bases 1 to 55707)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domano,M., Donegan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. 6
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This record contains 63 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 893: contig of 893 bp in length
* gap of unknown length

```

```

*      *      1785:  contig of 892 bp in length
*      *      gap of unknown length
*      *      2662:  contig of 877 bp in length
*      *      gap of unknown length
*      *      3517:  contig of 855 bp in length
*      *      gap of unknown length
*      *      3518:  contig of 751 bp in length
*      *      gap of unknown length
*      *      4269:  contig of 885 bp in length
*      *      gap of unknown length
*      *      5154:  contig of 904 bp in length
*      *      gap of unknown length
*      *      6058:  contig of 895 bp in length
*      *      gap of unknown length
*      *      6953:  contig of 863 bp in length
*      *      gap of unknown length
*      *      7816:  8695:  contig of 880 bp in length
*      *      gap of unknown length
*      *      8966:  9609:  contig of 914 bp in length
*      *      gap of unknown length
*      *      9610:  10502: contig of 893 bp in length
*      *      gap of unknown length
*      *      10503: 11392: contig of 890 bp in length
*      *      gap of unknown length
*      *      11393: 12274: contig of 882 bp in length
*      *      gap of unknown length
*      *      12275: 13030: contig of 756 bp in length
*      *      gap of unknown length
*      *      13031: 13909: contig of 879 bp in length
*      *      gap of unknown length
*      *      13910: 14768: contig of 859 bp in length
*      *      gap of unknown length
*      *      14769: 15644: contig of 876 bp in length
*      *      gap of unknown length
*      *      15645: 16566: contig of 922 bp in length
*      *      gap of unknown length
*      *      16567: 17446: contig of 880 bp in length
*      *      gap of unknown length
*      *      17447: 18339: contig of 893 bp in length
*      *      gap of unknown length
*      *      18340: 19236: contig of 897 bp in length
*      *      gap of unknown length
*      *      19237: 20155: contig of 919 bp in length
*      *      gap of unknown length
*      *      20156: 21081: contig of 926 bp in length
*      *      gap of unknown length
*      *      21082: 21949: contig of 868 bp in length
*      *      gap of unknown length
*      *      21950: 22879: contig of 930 bp in length
*      *      gap of unknown length
*      *      22880: 23733: contig of 854 bp in length
*      *      gap of unknown length
*      *      23734: 24608: contig of 875 bp in length
*      *      gap of unknown length
*      *      24609: 25491: contig of 883 bp in length
*      *      gap of unknown length
*      *      25492: 26378: contig of 887 bp in length
*      *      gap of unknown length
*      *      26379: 27257: contig of 879 bp in length
*      *      gap of unknown length
*      *      27258: 28144: contig of 887 bp in length
*      *      gap of unknown length
*      *      28145: 29040: contig of 896 bp in length
*      *      gap of unknown length
*      *      29041: 29904: contig of 864 bp in length
*      *      gap of unknown length
*      *      29905: 30787: contig of 883 bp in length
*      *      gap of unknown length
*      *      30788: 31661: contig of 874 bp in length
*      *      gap of unknown length
*      *      31662: 32594: contig of 933 bp in length
*      *      gap of unknown length
*      *      32595: 33479: contig of 885 bp in length

```

```

*      *      33480: 34403:  contig of 924 bp in length
*      *      gap of unknown length
*      *      34404: 35271:  contig of 868 bp in length
*      *      gap of unknown length
*      *      35272: 36161:  contig of 890 bp in length
*      *      gap of unknown length
*      *      36162: 37033:  contig of 872 bp in length
*      *      gap of unknown length
*      *      37034: 37925:  contig of 892 bp in length
*      *      gap of unknown length
*      *      37926: 38824:  contig of 899 bp in length
*      *      gap of unknown length
*      *      38825: 39688:  contig of 864 bp in length
*      *      gap of unknown length
*      *      39689: 40575:  contig of 887 bp in length
*      *      gap of unknown length
*      *      40576: 41466:  contig of 891 bp in length
*      *      gap of unknown length
*      *      41467: 42391:  contig of 925 bp in length
*      *      gap of unknown length
*      *      42392: 43266:  contig of 875 bp in length
*      *      gap of unknown length
*      *      43267: 44131:  contig of 865 bp in length
*      *      gap of unknown length
*      *      44132: 45006:  contig of 875 bp in length
*      *      gap of unknown length
*      *      45007: 45920:  contig of 914 bp in length
*      *      gap of unknown length
*      *      45921: 46801:  contig of 881 bp in length
*      *      gap of unknown length
*      *      46802: 47682:  contig of 881 bp in length
*      *      gap of unknown length
*      *      47683: 48573:  contig of 891 bp in length
*      *      gap of unknown length
*      *      48574: 49466:  contig of 893 bp in length
*      *      gap of unknown length
*      *      49467: 50377:  contig of 911 bp in length
*      *      gap of unknown length
*      *      50378: 51250:  contig of 873 bp in length
*      *      gap of unknown length
*      *      51251: 52137:  contig of 887 bp in length
*      *      gap of unknown length
*      *      52138: 53075:  contig of 938 bp in length
*      *      gap of unknown length
*      *      53076: 53931:  contig of 856 bp in length
*      *      gap of unknown length
*      *      53932: 54822:  contig of 891 bp in length
*      *      gap of unknown length
*      *      54823: 55707:  contig of 885 bp in length.
*      *      Location/Qualifiers
*      *      source
*      *      1. .55707
*      *      /organism="Homo sapiens"
*      *      /db_xref="taxon:9606"
*      *      /clone="115_K14"
*      *      /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT  16397 a 10660 c 10769 g 17182 t 699 others
ORIGIN

```

```

Query Match      13.1%; Score 18; DB 41; Length 55707;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CCACCGCACCTGGCTGAT 18
        |||
Db      51596 CCACCGCACCTGGCTGAT 51579

RESULT 30
AB026647      Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone: MJL12,
LOCUS      81542 bp      DNA      PLN      07-MAY-1999
DEFINITION

```

```

complete sequence.
AB026647
VERSION AB026647.1 GI:4757403
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MJL12
ORGANISM Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (sites)
Nakamura.K.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. II
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 81542)
AUTHORS Nakamura.Y.
TITLE Direct Submission
JOURNAL Submitted (28-APR-1999) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935(ex.7443),
Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     repeat_region    1..81542
                     /organism="Arabidopsis thaliana"
     repeat_region    /strain="Columbia"
                     /db_xref="taxon:3702"
     repeat_region    /chromosome="3"
                     /clone="MJL12"
     repeat_region    /clone_lib="Mitsui P1"
BASE COUNT 25859 a 14006 c 14834 g 26843 t
ORIGIN
Query Match      13.18; Score 18; DB 7; Length 81542;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 74 CCTTGAAGTCTTCCCAAGC 91
      |||||
Db 19284 CCTTGAAGTCTTCCCAAGC 19301

RESULT 31
HS931E15/c
LOCUS      HS931E15      81874 bp      DNA      PRI      23-NOV-1999
DEFINITION Human DNA sequence from clone 931E15 on chromosome Xg25. Contains
STSs, GSSs and genomic marker DXS8098, complete sequence.
ACCESSION AL0233575
VERSION    AL0233575.1 GI:3618163
KEYWORDS   HTG; DXS8098.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pavitt,R.
TITLE      Direct Submission
JOURNAL    Submitted (24-SEP-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT    On Sep 18, 1998 this sequence version replaced gi:3550203.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
931E15 is from the library RPC15 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:PCYFAC2
IMPORTANT: This sequence is not the entire insert of clone 931E15.
It may be shorter because we only sequence overlapping sections
once or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of 424J12 (282207) is at 36643 in this sequence.
The true right end of 506G2 (282213) is at 37431.
FEATURES             Location/Qualifiers
     source           1..81874
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q25"
                     /clone="RP5-931E15"
                     /clone_lib="RPCI-5"
     repeat_region    1..201
                     /note="Alusx repeat: matches 94..294 of consensus"
                     complement(228..531)
     repeat_region    /note="Alusx repeat: matches 292..1 of consensus"
                     569..687
     repeat_region    /note="AluJo repeat: matches 11..127 of consensus"
                     702..1002
     repeat_region    /note="Alusp repeat: matches 1..302 of consensus"
                     1010..1178
     repeat_region    /note="FRAM repeat: matches 5..166 of consensus"
                     complement(1365..1664)
     misc_feature     /note="match: STS G05254"
                     2391..2691
     repeat_region    /note="Alusx repeat: matches 1..300 of consensus"
                     2787..3067
     repeat_region    /note="AluJb repeat: matches 1..302 of consensus"
                     complement(4220..4537)
     repeat_region    /note="AluJb repeat: matches 300..1 of consensus"
                     4701..5004
     repeat_region    /note="AluJb repeat: matches 2..302 of consensus"
                     5170..5497
     repeat_region    /note="LI repeat: matches 3627..3966 of consensus"
                     5498..5643
     repeat_region    /note="Alusq repeat: matches 1..146 of consensus"
                     5644..5948
     repeat_region    /note="Aluy repeat: matches 1..301 of consensus"
                     5950..6115
     repeat_region    /note="Alusg repeat: matches 133..290 of consensus"
                     6156..6393
     repeat_region    /note="LI repeat: matches 3985..4239 of consensus"
                     6629..6979
     repeat_region    /note="LI repeat: matches 4377..4731 of consensus"
                     6978..7073
     repeat_region    /note="MER2 repeat: matches 1..95 of consensus"
                     complement(7078..7139)
     repeat_region    /note="MER2 repeat: matches 345..284 of consensus"
                     7238..7675
     repeat_region    /note="LI repeat: matches 4728..5161 of consensus"
                     7678..7969
     repeat_region    /note="Alusx repeat: matches 1..293 of consensus"
                     7970..8091
     repeat_region    /note="LI repeat: matches 5150..5270 of consensus"
                     8094..8396
     repeat_region    /note="Aluy repeat: matches 1..300 of consensus"
                     8400..8615
     repeat_region    /note="L1MB3 repeat: matches 47..276 of consensus"
                     8616..8733
     repeat_region    /note="AluJo repeat: matches 2..121 of consensus"
                     8743..8816
     repeat_region    /note="Aluy repeat: matches 228..301 of consensus"

```

```

repeat_region      8956..9146
/note="AluJo repeat: matches 2..300 of consensus"
repeat_region      9152..9362
/note="AluJo repeat: matches 87..299 of consensus"
repeat_region      9365..9667
/note="AluJo repeat: matches 1..302 of consensus"
repeat_region      9668..10289
/note="L1MA10 repeat: matches 297..919 of consensus"
/note="AluSx repeat: matches 302..3 of consensus"
10710..11004
/note="AluSx repeat: matches 7..301 of consensus"
11299..11939
/note="L1PA8 repeat: matches 271..910 of consensus"
11983..12087
/note="L1MB7 repeat: matches 6..109 of consensus"
12100..12401
/note="AluSx repeat: matches 4..302 of consensus"
12439..12940
/note="L1ME3 repeat: matches 166..685 of consensus"
/note="match: STS L40936"
13604..13904
/note="AluSx repeat: matches 2..302 of consensus"
13958..14050
/note="AluSg repeat: matches 1..93 of consensus"
14055..14323
/note="AluJb repeat: matches 60..301 of consensus"
14514..14836
/note="AluJb repeat: matches 13..302 of consensus"
15126..15428
/note="AluSx repeat: matches 1..302 of consensus"
15626..15730
/note="AluJb repeat: matches 196..300 of consensus"
16118..16420
/note="AluJb repeat: matches 1..290 of consensus"
/note="FLAM_A repeat: matches 117..6 of consensus"
complement(17160..17224)
/note="MIR repeat: matches 149..84 of consensus"
18120..18226
/note="match: STS L24586"
complement(18124..18827)
/note="match: GSSs B16853 B14990"
complement(19378..19676)
/note="AluSp repeat: matches 303..2 of consensus"
19901..20160
/note="AluY repeat: matches 39..298 of consensus"
complement(20716..20882)
/note="MIR repeat: matches 260..69 of consensus"
complement(20824..21385)
/note="match: GSSs B94184 A0010265"
21550..23628
/note="putative CpG island"
complement(22542..22581)
/note="MIR2 repeat: matches 134..95 of consensus"
complement(23094..24185)
/note="SVA repeat: matches 1101..1 of consensus"
24186..24231
/note="23 copies 2 mer aa 80% conserved"
24520..24884
/note="L1PA5 repeat: matches 517..890 of consensus"
complement(25504..25687)
/note="MER20 repeat: matches 155..1 of consensus"
26421..26484
/note="32 copies 2 mer tt 83% conserved"
complement(26486..26661)
/note="AluJb repeat: matches 195..5 of consensus"
26867..27164
/note="AluY repeat: matches 3..300 of consensus"
27165..27462
/note="AluSp repeat: matches 2..302 of consensus"
28080..28392

```

```

repeat_region      /note="AluSx repeat: matches 1..298 of consensus"
complement(28817..29089)
repeat_region      /note="AluSx repeat: matches 298..2 of consensus"
29122..29252
/note="L1 repeat: matches 3151..3280 of consensus"
complement(29957..30200)
/note="MLT2_internal repeat: matches 5375..5124 of
consensus"
complement(30513..30706)
/note="MLT2_internal repeat: matches 4754..4572 of
consensus"
30728..31025
/note="AluSx repeat: matches 1..299 of consensus"
31059..31356
/note="AluSg repeat: matches 1..296 of consensus"
31359..31505
/note="MER4B repeat: matches 11..148 of consensus"
32041..32319
/note="AluSx repeat: matches 1..287 of consensus"
complement(32381..33599)
/note="MLT2_internal repeat: matches 4572..3316 of
consensus"
33603..33906
/note="AluJo repeat: matches 1..302 of consensus"
34020..34321
/note="AluJb repeat: matches 1..301 of consensus"
complement(34388..34933)

Query Match      13.1%; Score 18; DB 11; Length 81874;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 4714 CCACCGCACCTGGCTGAT 4697

RESULT 32
AC004472/c
LOCUS AC004472 86765 bp DNA PRI 25-MAR-1998
DEFINITION Homo sapiens chromosome 9, p1 clone 11659, complete sequence.
AC004472
VERSION AC004472.1 GI:2984582
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 86765)
AUTHORS Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Ganes,J., Danganan,L., Poundstone,P.,
Christensen,M., Georgescu,A., Avila,J., Liu,S., Attix,C.,
Andreise,T., Frankheim,M., Anico-Keller,G., Coefield,J., Duarte,S.,
Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A.,
Montgomery,M., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.O.
and Carrano,A.V.
TITLE Sequence analysis of a human p1 clone containing the XRCC9 DNA
repair gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 86765)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
FEATURES
source
1..86765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="P1.11659"
/chromosome="9"
/map="9p13"
/tissue_type="Fibroblast or foreskin"

```





```

repeat_region 26514..26804
/rpt_family="AluJb"
complement(26831..27035)
/rpt_family="MLT1D"
complement(27042..27333)
/rpt_family="AluY"
complement(27334..27637)
/rpt_family="AluSp"
complement(27644..27716)
/rpt_family="MLT1D"
27727..28020
/rpt_family="AluSx"
28861..28972
/rpt_family="AluSx"
29177..29471
/rpt_family="7SLRNA"
complement(29214..29339)
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 50.000"
complement(29628..29915)

Query Match 13.1%; Score 18; DB 11; Length 86765;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 19053 CCACCGCACCTGGCTGAT 19036

RESULT 33
HS398C22 89328 bp DNA PRI 23-NOV-1999
LOCUS Homo sapiens DNA sequence from PAC 398C22 on chromosome
DEFINITION 2q11.2-qter. Contains Brain Protein E46 like sequences, ESTs.
ACCESSION 293784
VERSION 293784.1
KEYWORDS 2q11.2-qter; Brain Protein E46-like.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89328)
Burgess,J.
Direct Submission
Submitted (31-JUL-1997) Chromosome 22 Project Group
(http://www.sanger.ac.uk/HGP/Chr22/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquerry@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 7, 1997 this sequence version replaced gi:1929050.
IMPORTANT: This sequence is not the entire insert of clone 398C22.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre chromosome 22
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true right end of clone 941f9 is at 104.
The true right end of clone 398C22 is at 89328.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
398C22 is from the library RPCI3 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong.
For further details see http://bacpac.med.buffalo.edu/.

FEATURES
Location/Qualifiers
source
1..89328
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2-qter"
/clone="Rp3-398C22"
/clone_lib="RPCI-3"
511..575
repeat_region
/notes="AluJb repeat: matches 296..232 of consensus;
incomplete repeat"
577..735
repeat_region
/notes="AluJb repeat: matches 153..3 of consensus;
incomplete repeat"
2291..2582
repeat_region
/notes="AluJb repeat: matches 300..2 of consensus"
3011..3311
repeat_region
/notes="AluSx repeat: matches 1..301 of consensus"
3624..3763
repeat_region
/notes="L1 repeat: matches 5042..4898 of consensus"
3924..4226
repeat_region
/notes="AluJo repeat: matches 302..1 of consensus"
6512..6809
repeat_region
/notes="AluSg repeat: matches 1..297 of consensus"
7753..8041
repeat_region
/notes="AluSg repeat: matches 1..290 of consensus"
9331..9533
repeat_region
/notes="MIR repeat: matches 38..251 of consensus"
9687..96512
gene
/genes="dj398C22.1"
join(<9687..9877,12970..13052,20256..20352,22663..22821,
38387..38467,49399..49564,58705..58813,60343..60512)
/genes="dj398C22.1"
/notes="E46-like contains exons 2-9 continues in 284478"
/codon_start=1
/product="dj398C22.1 (novel protein, ortholog of mouse
brain protein E46)"
/protein_id="CAB07856.1"
/db_xref="GI:2956661"
/db_xref="SPTREMBL:O14998"
/translation="ETAPRTIFQRVLDILKSHAVELACRDPQSQVENLASSIQLITE
CFCLRNACIECVSNQNSIRNLDITIGVAVDLILFLRELVEQSLITAFRCQFLGN
IASRNEDSQSIYVHAFPELFLSCLNHPDKKIYVSSMILFTSLNHERKLEFLNI
AIDVIDAYQKHPESEWPFLLITDILFLKSPELVQAMPKLNQERVTLIDLMIAKITSD
EPLTKDDIPELRLHAEILASTVDQCKTVLKASEPPDDEALATIRLDVCEVTY
NTELLGLVQVFCGLLERVIDLLRLVHVACKETNIFSNCGCVRAEGDISNVANGFKSH
LRLIGNLCYKKNRNDQK"
9965..10269
repeat_region
/notes="AluJb repeat: matches 1..302 of consensus"
11343..11529
repeat_region
/notes="MIR repeat: matches 223..30 of consensus"
11588..11877
repeat_region
/notes="AluSg repeat: matches 1..291 of consensus"
11942..12099
repeat_region
/notes="MIR repeat: matches 86..259 of consensus"
13267..13632
repeat_region
/notes="MLT1A1 repeat: matches 358..1 of consensus"
13635..13927
repeat_region
/notes="MLT1B repeat: matches 44..358 of consensus"
13947..14046
repeat_region
/notes="MIR2 repeat: matches 109..15 of consensus"
14344..14473
repeat_region
/notes="AluJb repeat: matches 1..131 of consensus;
incomplete repeat"
14474..14770
repeat_region
/notes="AluSp repeat: matches 1..297 of consensus"
14771..14934
repeat_region
/notes="AluSx repeat: matches 138..302 of consensus;
incomplete repeat"
15023..15315
repeat_region
/notes="AluY repeat: matches 1..298 of consensus"

```

```
repeat_region 15455..15765
/Note="AluSg repeat: matches 297. .1 of consensus"
prim_transcript 16256..17096
/Genes="dJ398C22.1"
/Note="match: multiple ESTs; match: R10652 R99219 R99220
R10653"
repeat_region 17008..17105
/Note="L1MB4 repeat: matches 904. .809 of consensus"
repeat_region 17116..17408
/Note="AluSg repeat: matches 294. .1 of consensus"
repeat_region 17414..17602
/Note="L1MB5 repeat: matches 795. .608 of consensus"
repeat_region 17605..17903
/Note="AluY repeat: matches 300. .1 of consensus"
repeat_region 17905..18486
/Note="L1MB8 repeat: matches 622. .4 of consensus"
repeat_region 18489..18603
/Note="L1 repeat: matches 4584. .4471 of consensus"
repeat_region 19455..19765
/Note="AluSg repeat: matches 301. .3 of consensus"
repeat_region 19822..20085
/Note="AluSx repeat: matches 299. .1 of consensus"
repeat_region 20399..20661
/Note="L1MB2 repeat: matches 654. .911 of consensus"
repeat_region 20818..20913
/Note="3 copies of 32 mer 85 & conserved"
repeat_region 24126..25411
/Note="L1 repeat: matches 4782. .3450 of consensus"
repeat_region 25415..25707
/Note="AluSg repeat: matches 291. .1 of consensus"
repeat_region 25712..25797
/Note="L1 repeat: matches 3463. .3375 of consensus"
repeat_region 25800..26146
/Note="MER1B repeat: matches 1. .337 of consensus"
repeat_region 26148..27122
/Note="L1 repeat: matches 3379. .2343 of consensus"
repeat_region 28621..28908
/Note="AluSx repeat: matches 293. .5 of consensus"
repeat_region 30498..30588
/Note="MIR2 repeat: matches 1. .96 of consensus"
repeat_region 31144..31683
/Note="L1MB2 repeat: matches 545. .1088 of consensus"
repeat_region 33748..34075
/Note="L1MB1 repeat: matches 633. .290 of consensus"
repeat_region 34358..34655
/Note="AluJo repeat: matches 301. .1 of consensus"
repeat_region 35400..35616
/Note="MER30 repeat: matches 230. .1 of consensus"
repeat_region 35953..36259
/Note="AluJo repeat: matches 2. .302 of consensus"
repeat_region 36300..36610
/Note="AluSx repeat: matches 301. .9 of consensus"
repeat_region 39850..39964
/Note="MIR2 repeat: matches 5. .125 of consensus"
repeat_region 40247..40290
/Note="L1MB3 repeat: matches 923. .880 of consensus"
repeat_region 40291..40470
/Note="FAM repeat: matches 174. .5 of consensus"
repeat_region 40475..40598
/Note="L1MB2 repeat: matches 880. .757 of consensus"
repeat_region 40597..40744
/Note="L1MB8 repeat: matches 176. .314 of consensus"
repeat_region 40744..40918
/Note="MSTB repeat: matches 394. .217 of consensus"
repeat_region 41353..41434
/Note="MIR repeat: matches 145. .234 of consensus"
repeat_region 41694..41761
/Note="MIR repeat: matches 191. .119 of consensus"
repeat_region 41762..41907
/Note="AluSg repeat: matches 154. .299 of consensus;
incomplete repeat"
repeat_region 43244..43313
/Note="2 copies of 35 mer 99 & conserved"
```

```
repeat_region 43315..43384
/Note="2 copies of 35 mer 91 & conserved"
repeat_region 43385..43649
/Note="L1PB2 repeat: matches 874. .604 of consensus"
repeat_region 43650..43944
/Note="AluSx repeat: matches 3. .301 of consensus"
repeat_region 43949..44542
/Note="L1PB1 repeat: matches 608. .1 of consensus"
repeat_region 44395..44659
/Note="L1 repeat: matches 5390. .3097 of consensus"
repeat_region 46669..46954
/Note="AluJb repeat: matches 293. .1 of consensus"
repeat_region 46958..47286
Query Match 13.1%; Score 18; DB 11; Length 89328;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18
|||||
Db 34505 CCACCGCACCTGGCTGAT 34522

RESULT 34
AC002126 89818 bp DNA 15-AUG-1997
LOCUS Homo sapiens DNA from chromosome 19-cosmids R30102:R29350:R27740
DEFINITION containing MEF2B, genomic sequence, complete sequence.
ACCESSION AC002126
VERSION AC002126.1 GI:2329908
KEYWORDS Mtg.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 89818)
AUTHORS Lamerdin,J.E., Kyle,A., Elliot,J., McCready,P.M., Kobayashi,A.,
Ow,D., Carrano,A.V. and Garcia,E.
Direct Submission
Submitted (15-AUG-1997) Biology and Biotechnology Research Program,
Lawrence Livermore National Laboratory, 7000 East Ave, Livermore,
CA 94550, USA
REMARK Genomic Structure of the Human MEF2B Gene
COMMENT Consensus sequence assembled from overlapping cosmids R30102 (bases
1- 36,529), R29350 (bases 29,365- 71,464), and R27740 (bases
54,941- 89,818).
FEATURES
Location/Qualifiers
source 1..89818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p12"
/chromosome="19"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/Note="LL19NC03 library constructed at LLNL from
flow-sorted chromosomes from Hybrid 5HL2-B, which carries
chromosome 19 as its only human chromosome"
complement(177..452)
/rpt_family="ALU"
800..1089
/rpt_family="ALU"
repeat_region
repeat_region
CDS
<1304..1611
/Note="hypothetical partial orf most similar to orf from
C. elegans cosmid F43G9.3 (279755); putative exon, Xgrail
1.3c, excellent"
/codon_start=3
/product="R31240_1"
```

misc_feature	/protein_id="AAB86983.1" /db_xref="GI:2329910" translation="YSGRRQYPERMIFGACAGLIGSASYPDYVRRRQWAGVTG IPRASTARTLTIVREEGAVKGLKGSMMNVKGPVAVGSFTFDLMQLLRHLQS" 2171..2405 /note="BLASTN similarity to H67311 (1..236); match: 1, sapiens cDNA clone 229611 5'." score: 4.1e-83; database searched: est; yu51b02.r1 Homo complement(2403..2725) misc_feature	/note="BLASTN similarity to H03990 (387..426); match: 0.9, sapiens cDNA clone 151863 3'." complement(3243..3284) /note="BLASTN similarity to H03990 (347..388); match: 0.95, score: 3.4e-142; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3255..3284) misc_feature	/note="BLASTN similarity to R10902 (336..365); match: 0.76, score: 2.1e-117; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3266..3284) misc_feature	/note="BLASTN similarity to R22559 (356..374); match: 0.94, score: 1.1e-108; database searched: est; yh23hl2.s1 Homo sapiens cDNA clone 130631 3'.>gb G30262 G30262 human STS SHGC-36640" complement(3270..3318) misc_feature	/note="BLASTN similarity to R10902 (301..349); match: 0.85, score: 2.8e-125; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3281..3318) misc_feature	/note="BLASTN similarity to H03990 (312..349); match: 0.73, score: 3.1e-97; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3297..3432) misc_feature	/note="BLASTN similarity to H03990 (197..332); match: 0.97, score: 3.4e-142; database searched: est; yj47b08.s1 Homo sapiens cDNA clone 151863 3'." complement(3297..3321) misc_feature	/note="BLASTN similarity to R48002 (306..330); match: 0.92, score: 1.1e-95; database searched: est; yj63g05.s1 Homo sapiens cDNA clone 153464 3'." complement(3298..3621) misc_feature	/note="BLASTN similarity to Z39994 (1..324); match: 0.99, score: 6.2e-126; database searched: est; H. sapiens partial cDNA sequence" complement(3310..3362) misc_feature	/note="BLASTN similarity to R22559 (272..324); match: 0.83, score: 1.1e-108; database searched: est; yh23hl2.s1 Homo sapiens cDNA clone 130631 3'.>gb G30262 G30262 human STS SHGC-36640" 3310..3339 misc_feature	/note="BLASTN similarity to R48116 (353..382); match: 0.8, score: 2.0e-134; database searched: est; yj63g05.r1 Homo sapiens cDNA clone 153464 5'." complement(3310..3432) misc_feature	/note="BLASTN similarity to R10902 (186..308); match: 0.97, score: 2.8e-125; database searched: est; yf38f01.s1 Homo sapiens cDNA clone 129145 3'." complement(3322..3362) misc_feature	/note="BLASTN similarity to R48002 (262..302); match: 0.97, score: 1.1e-95; database searched: est; yj63g05.s1 Homo sapiens cDNA clone 153464 3'." 3332..3397 misc_feature	/note="BLASTN similarity to R48116 (379..444); match: 0.74, score: 2.0e-134; database searched: est; yj63g05.r1 Homo sapiens cDNA clone 153464 5'." complement(3336..3432) misc_feature	/note="BLASTN similarity to H68521 (192..288); match: 0.97, score: 2.9e-106; database searched: est; yu51b02.s1 Homo sapiens cDNA clone 229611 3'." complement(3339..3391) misc_feature	/note="BLASTN similarity to R22559 (242..294); match: 0.66, score: 5.6e-102; database searched: est; yh23hl2.s1 Homo sapiens cDNA clone 130631 3'.>gb G30262 G30262 human STS SHGC-36640" complement(3371..3624) misc_feature	/note="BLASTN similarity to Z16003 (1..254); match: 0.95, score: 3.5e-89; database searched: est; H. sapiens partial cDNA sequence, clone 56B12" complement(3375..3629) misc_feature	/note="BLASTN similarity to R22559 (2..256); match: 0.98, score: 1.1e-108; database searched: est; yh23hl2.s1 Homo
--------------	--	---	---	--	---	--	---	--	--	--	--	---	--	---	---	--	--	---



```

* 51004 52956: contig of 1953 bp in length
* gap of unknown length
* 52957 54144: contig of 1188 bp in length
* gap of unknown length
* 54145 55811: contig of 1667 bp in length
* gap of unknown length
* 55812 56739: contig of 928 bp in length
* gap of unknown length
* 56740 58330: contig of 1591 bp in length
* gap of unknown length
* 58331 60299: contig of 1969 bp in length
* gap of unknown length
* 60300 62534: contig of 2235 bp in length
* gap of unknown length
* 62535 64966: contig of 2432 bp in length
* gap of unknown length
* 64967 68332: contig of 3366 bp in length
* gap of unknown length
* 68333 93841: contig of 25509 bp in length.
FEATURES
Source
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="PL-360B11"
BASE COUNT 22734 a 23592 c 24325 g 23115 t 75 others
ORIGIN
|||||

Query Match 13.1%; Score 18; DB 41; Length 93841;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 CTAAGTGAATCTTCT 76
Db 23119 CTAAGTGAATCTTCT 23102

RESULT 36
HSJ506G21/c 95425 bp DNA HTG 08-NOV-1996
LOCUS Homo sapiens chromosome X clone J506G21, *** SEQUENCING IN PROGRESS
DEFINITION
***, in unordered pieces.
282213
ACCESSION 282213.1 GI:1666405
VERSION HTG; HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 95425)
AUTHORS Unknown.
TITLE Direct Submission
JOURNAL Submitted (06-NOV-1996) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT Order of segments is not known; 800 n's separate segments.
Unfinished sequence: dj506G2 Contig_ID: 01828 Length: 797 bp
Unfinished sequence: dj506G2 Contig_ID: 01922 Length: 37027 bp
Unfinished sequence: dj506G2 Contig_ID: 01957 Length: 1258 bp
Unfinished sequence: dj506G2 Contig_ID: 00779 Length: 53943 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Source
Location/Qualifiers
1..95425
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="J506G21"
BASE COUNT 30849 a 19403 c 17797 g 24967 t 2409 others
ORIGIN
|||||

```

```

Query Match 13.1%; Score 18; DB 32; Length 95425;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGACCTGGCTGAT 18
Db 5906 CCACCGACCTGGCTGAT 5889

RESULT 37
AC011336 96099 bp DNA HTG 06-OCT-1999
LOCUS Homo sapiens chromosome 5 clone CIT-HSPC_250I13, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
AC011336
ACCESSION AC011336.1 GI:6013604
VERSION HTG; HTGS_PHASE1.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 96099)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 96099)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 705: contig of 705 bp in length
* gap of unknown length
* 706 1367: contig of 662 bp in length
* gap of unknown length
* 1368 2143: contig of 776 bp in length
* gap of unknown length
* 2144 2940: contig of 797 bp in length
* gap of unknown length
* 2941 4813: contig of 1873 bp in length
* gap of unknown length
* 4814 7980: contig of 3167 bp in length
* gap of unknown length
* 7981 49500: contig of 41520 bp in length
* gap of unknown length
* 49501 96099: contig of 46399 bp in length.
FEATURES
Source
Location/Qualifiers
1..96099
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CIT-HSPC_250I13"
BASE COUNT 23213 a 24003 c 24571 g 24303 t 9 others
ORIGIN
|||||

Query Match 13.1%; Score 18; DB 41; Length 96099;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 CTAAGTGAATCTTCT 76
Db 68149 CTAAGTGAATCTTCT 68166

```

## RESULT 38

AC013370 100091 bp DNA HTG 20-NOV-1999  
 LOCUS Homo sapiens clone RP11-12D16, LOW-PASS SEQUENCE SAMPLING.  
 DEFINITION AC013370  
 ACCESSION AC013370.3 GI:6456269  
 VERSION  
 KEYWORDS HTG; HTGS-PHASED.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 100091)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

## TITLE

Unpublished

## JOURNAL

2 (bases 1 to 100091)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,  
 Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Horton,L.,  
 Galagan,J., Gardy,J.S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lehoczy,J., Lieue,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 20, 1999 this sequence version replaced gi:6425738.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997).

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Project Information

Center project name: L3253

Center clone name: L2\_D\_16

-----

\* NOTE: This record contains 110 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved. 910: contig of 910 bp in length

\* 1 gap of unknown length

\* 911 1846: contig of 936 bp in length

\* gap of unknown length

\* 1847 2768: contig of 922 bp in length

\* gap of unknown length

\* 2769 3751: contig of 983 bp in length

\* gap of unknown length

\* 3752 4608: contig of 857 bp in length

\* gap of unknown length

\* 4609 5507: contig of 899 bp in length

\* gap of unknown length

\* 5508 6398: contig of 891 bp in length

\* gap of unknown length

\* 6399 7310: contig of 912 bp in length

\* gap of unknown length

\* 7311 8239: contig of 929 bp in length

\* gap of unknown length

\* 8240 9173: contig of 934 bp in length

\* 9174 10107: contig of 934 bp in length

\* gap of unknown length

\* 10108 11014: contig of 907 bp in length

\* gap of unknown length

\* 11015 11934: contig of 920 bp in length

\* gap of unknown length

\* 11935 12855: contig of 921 bp in length

\* gap of unknown length

\* 12856 13746: contig of 891 bp in length

\* gap of unknown length

\* 13747 14742: contig of 996 bp in length

\* gap of unknown length

\* 14743 15668: contig of 926 bp in length

\* gap of unknown length

\* 15669 16623: contig of 955 bp in length

\* gap of unknown length

\* 16624 17547: contig of 924 bp in length

\* gap of unknown length

\* 17548 18486: contig of 939 bp in length

\* gap of unknown length

\* 18487 19433: contig of 947 bp in length

\* gap of unknown length

\* 19434 20378: contig of 945 bp in length

\* gap of unknown length

\* 20379 21271: contig of 893 bp in length

\* gap of unknown length

\* 21272 22176: contig of 905 bp in length

\* gap of unknown length

\* 22177 23090: contig of 914 bp in length

\* gap of unknown length

\* 23091 24002: contig of 912 bp in length

\* gap of unknown length

\* 24003 24930: contig of 928 bp in length

\* gap of unknown length

\* 24931 25870: contig of 940 bp in length

\* gap of unknown length

\* 25871 26812: contig of 942 bp in length

\* gap of unknown length

\* 26813 27775: contig of 963 bp in length

\* gap of unknown length

\* 27776 28688: contig of 913 bp in length

\* gap of unknown length

\* 28689 29669: contig of 981 bp in length

\* gap of unknown length

\* 29670 30581: contig of 912 bp in length

\* gap of unknown length

\* 30582 31525: contig of 944 bp in length

\* gap of unknown length

\* 31526 32441: contig of 916 bp in length

\* gap of unknown length

\* 32442 33365: contig of 924 bp in length

\* gap of unknown length

\* 33366 34290: contig of 925 bp in length

\* gap of unknown length

\* 34291 35205: contig of 915 bp in length

\* gap of unknown length

\* 35206 36136: contig of 931 bp in length

\* gap of unknown length

\* 36137 37079: contig of 943 bp in length

\* gap of unknown length

\* 37080 38018: contig of 939 bp in length

\* gap of unknown length

\* 38019 38963: contig of 945 bp in length

\* gap of unknown length

\* 38964 39855: contig of 892 bp in length

\* gap of unknown length

\* 39856 40787: contig of 932 bp in length

\* gap of unknown length

\* 40788 41729: contig of 942 bp in length

\* gap of unknown length

\* 41730 42618: contig of 889 bp in length

\* gap of unknown length

*	42619	43516:	contig of 898 bp in length
*			gap of unknown length
*	43517	44467:	contig of 951 bp in length
*			gap of unknown length
*	44468	45413:	contig of 946 bp in length
*			gap of unknown length
*	45414	46346:	contig of 933 bp in length
*			gap of unknown length
*	46347	47264:	contig of 918 bp in length
*			gap of unknown length
*	47265	48176:	contig of 912 bp in length
*			gap of unknown length
*	48177	49115:	contig of 939 bp in length
*			gap of unknown length
*	49116	50009:	contig of 894 bp in length
*			gap of unknown length
*	50010	50946:	contig of 937 bp in length
*			gap of unknown length
*	50947	51874:	contig of 928 bp in length
*			gap of unknown length
*	51875	52808:	contig of 934 bp in length
*			gap of unknown length
*	52809	53724:	contig of 916 bp in length
*			gap of unknown length
*	53725	54679:	contig of 955 bp in length
*			gap of unknown length
*	54680	55618:	contig of 939 bp in length
*			gap of unknown length
*	55619	56533:	contig of 915 bp in length
*			gap of unknown length
*	56534	57470:	contig of 937 bp in length
*			gap of unknown length
*	57471	58390:	contig of 920 bp in length
*			gap of unknown length
*	58391	59321:	contig of 931 bp in length
*			gap of unknown length
*	59322	60227:	contig of 906 bp in length
*			gap of unknown length
*	60228	61197:	contig of 970 bp in length
*			gap of unknown length
*	61198	62117:	contig of 920 bp in length
*			gap of unknown length
*	62118	63053:	contig of 936 bp in length
*			gap of unknown length
*	63054	63957:	contig of 904 bp in length
*			gap of unknown length
*	63958	64910:	contig of 953 bp in length
*			gap of unknown length
*	64911	65827:	contig of 917 bp in length
*			gap of unknown length
*	65828	66801:	contig of 974 bp in length
*			gap of unknown length
*	66802	67714:	contig of 913 bp in length
*			gap of unknown length
*	67715	68647:	contig of 933 bp in length
*			gap of unknown length
*	68648	69574:	contig of 927 bp in length
*			gap of unknown length

Neurotactin, Myotendinous antigen)-LIKE gene and a mitochondrial/chloroplast 30S ribosomal protein S14-LIKE gene preceded by a CpG island. Contains ESTs, genomic marker D1S2691 and STSS.

259297  
259297.1 GI:2955179  
Cytotactin; Hexabrachion; mitochondrial 30S ribosomal; Myotendinous antigen; Neurotactin; S14; Tenascin.

human.

SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 100997)  
Direct Submission  
Pavitt,R.  
JOURNAL Submitted (16-MAR-1998) sanger.ac.uk/HGP/Chr11/ Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Clone requests: clonerequests@sanger.ac.uk  
On Mar 28, 1998 this sequence version replaced gi:2578077.  
IMPORTANT: This sequence is not the entire insert of clone 262D12. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

COMMENT  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above.  
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre chromosome 1 mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1/>  
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
262D12 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.  
For further details see <http://bacpac.med.buffalo.edu/> The true left end of clone 102620 is at 100895 in this sequence. The true left end of clone 262D12 is at 1 in this sequence. The true right end of clone 1141019 is at 50699 in this sequence.

FEATURES  
Source  
1..100997  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="q23.3-24.3"  
/clone="Rpl-262D12"  
/clone\_lib="RPCI-1"  
708..1270  
/note="MER25 repeat: matches 1556..2130 of consensus"  
1723..1915  
/note="L1 repeat: matches 1993..2184 of consensus"  
1919..2224  
/note="AluX repeat: matches 1..302 of consensus"  
2308..2495  
/note="L1 repeat: matches 2239..2425 of consensus"  
2498..3316  
/note="L1 repeat: matches 3786..4615 of consensus"  
3334..3502  
/note="AluX repeat: matches 302..134 of consensus; incomplete repeat"  
3503..3804  
/note="AluX repeat: matches 302..1 of consensus"  
3815..3947  
/note="AluS repeat: matches 133..1 of consensus; incomplete repeat"  
3948..4733  
/note="L1 repeat: matches 4617..5390 of consensus"





REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 109210)

Blakey, S.

Direct Submission

Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequests@sanger.ac.uk

On Aug 18, 1999 this sequence version replaced gi:5708178.

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information  
on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence is  
the entire insert of clone 863C7. This sequence has been finished  
according to sequence map criteria as follows. An attempt is made  
to resolve all sequencing problems, such as compressions and  
repeats, but not necessarily within known annotated human repeat  
sequence elements (e.g. Alu). Where the sequence is ambiguous,  
there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of  
human chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr20>

863C7 is from the library RPCI-5 constructed at the Roswell Park  
Cancer Institute by the group of Pieter de Jong. For further  
details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

## FEATURES

source

1..109210  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="p12.3-13"  
/clone\_lib="RPCI-5"  
/clone="RP5-863C7"

BASE COUNT 32132 a 23899 c 22881 g 30298 t

ORIGIN

Query Match 13.1%; Score 18; DB 11; Length 109210;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 99544 CCACCGCACCTGGCTGAT 99527

RESULT 41

HS112K5/c

LOCUS 109891 bp DNA PRI 23-NOV-1999

DEFINITION Human DNA sequence from clone 112K5 on chromosome Xp11, complete  
sequence.

ACCESSION 285987.13 GI:5531490

KEYWORDS HTG.

SOURCE human.

ORGANISM  
Homo sapiensEukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 109891)

Pearce, A.

Direct Submission

Submitted (19-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequests@sanger.ac.uk

On Jul 19, 1999 this sequence version replaced gi:5459255.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL

This sequence is the entire insert of clone 112K5. This sequence  
has been finished according to sequence map criteria as follows. An  
attempt is made to resolve all sequencing problems, such as  
compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

This sequence was generated from part of bacterial clone contigs of  
human chromosome X, constructed by the Sanger Centre Chromosome X  
Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/ChrX>

112K5 is from the library RPCI1 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

## FEATURES

source

1..109891  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="p11"  
/clone="RP1-112K5"  
/clone\_lib="RPCI-1"

BASE COUNT 29357 a 25957 c 26406 g 28171 t

ORIGIN

Query Match 13.1%; Score 18; DB 10; Length 109891;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 78273 CCACCGCACCTGGCTGAT 78256

RESULT 42

AC010679.2

WPCOMMENT

Sequence split into 4 fragments LOCUS AC010679 Accession AC010679

Fragment Name

Begin End

AC010679\_0 1 110000

AC010679\_1 100001 210000

AC010679\_2 200001 310000

AC010679\_3 300001 350997

Continuation (3 of 4) of AC010679 from base 200001 (AC010679 Homo sapiens clone NH012

Query Match 13.1%; Score 18; DB 41; Length 110000;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCACCGCACCTGGCTGAT 18

|||||

Db 78298 CCACCGCACCTGGCTGAT 78315

RESULT 43

AC015845\_0/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC015845 Accession AC015845

Fragment Name

Begin End

AC015845\_0 1 110000

AC015845\_1 100001 210000

AC015845\_2 200001 310000

AC015845\_3 300001 410000

AC015845\_4 400001 510000

```

AC015845.5      500001      573550
LOCUS      AC015845      573550 bp      DNA      HTG      17-NOV-1999
DEFINITION      Homo sapiens chromosome 17 clone RP11-343K8 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION      AC015845
VERSION      AC015845.1      GI:6446823
KEYWORDS      HTG: HTGS_PHASE0: NULL.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 573550)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckler,R., Boguslavskiy,L., Bouckhgalter,B.,
            Brown,A., Castle,A., Collangelo,M., Collins,S., Collamore,A.,
            Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hayes,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RN/RepeatMasker.html
            ----- Project Information
            Center project name: L479
            Center clone name: 343_K_8
            -----
            * NOTE: This record contains 687 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            1      780:      contig of 780 bp in length
            *      gap of unknown length
            *      781      1566:      contig of 786 bp in length
            *      gap of unknown length
            *      1567      2334:      contig of 768 bp in length
            *      gap of unknown length
            *      2335      3108:      contig of 774 bp in length
            *      gap of unknown length
            *      3109      3900:      contig of 792 bp in length
            *      gap of unknown length
            *      3901      4700:      contig of 800 bp in length
            *      gap of unknown length
            *      4701      5482:      contig of 782 bp in length
            *      gap of unknown length
            *      5483      6257:      contig of 775 bp in length
            *      gap of unknown length
            *      6258      7034:      contig of 777 bp in length
            *      gap of unknown length
            *      7035      7795:      contig of 761 bp in length
            *      gap of unknown length
            *      7796      8572:      contig of 777 bp in length
            *      gap of unknown length
            *
            8573      9355:      contig of 783 bp in length
            *      gap of unknown length
            *      9356      10127:      contig of 772 bp in length
            *      gap of unknown length
            *      10128      10892:      contig of 765 bp in length
            *      gap of unknown length
            *      10893      11643:      contig of 751 bp in length
            *      gap of unknown length
            *      11644      12435:      contig of 792 bp in length
            *      gap of unknown length
            *      12436      13236:      contig of 801 bp in length
            *      gap of unknown length
            *      13237      14022:      contig of 786 bp in length
            *      gap of unknown length
            *      14023      14808:      contig of 786 bp in length
            *      gap of unknown length
            *      14809      15592:      contig of 784 bp in length
            *      gap of unknown length
            *      15593      16375:      contig of 783 bp in length
            *      gap of unknown length
            *      16376      17118:      contig of 743 bp in length
            *      gap of unknown length
            *      17119      17909:      contig of 791 bp in length
            *      gap of unknown length
            *      17910      18698:      contig of 789 bp in length
            *      gap of unknown length
            *      18699      19488:      contig of 790 bp in length
            *      gap of unknown length
            *      19489      20168:      contig of 680 bp in length
            *      gap of unknown length
            *      20169      20951:      contig of 783 bp in length
            *      gap of unknown length
            *      20952      21731:      contig of 780 bp in length
            *      gap of unknown length
            *      21732      22519:      contig of 788 bp in length
            *      gap of unknown length
            *      22520      23304:      contig of 785 bp in length
            *      gap of unknown length
            *      23305      24077:      contig of 773 bp in length
            *      gap of unknown length
            *      24078      24868:      contig of 791 bp in length
            *      gap of unknown length
            *      24869      25668:      contig of 800 bp in length
            *      gap of unknown length
            *      25669      26451:      contig of 783 bp in length
            *      gap of unknown length
            *      26452      27234:      contig of 783 bp in length
            *      gap of unknown length
            *      27235      27992:      contig of 758 bp in length
            *      gap of unknown length
            *      27993      28776:      contig of 784 bp in length
            *      gap of unknown length
            *      28777      29566:      contig of 790 bp in length
            *      gap of unknown length
            *      29567      30372:      contig of 806 bp in length
            *      gap of unknown length
            *      30373      31172:      contig of 800 bp in length
            *      gap of unknown length
            *      31173      31959:      contig of 787 bp in length
            *      gap of unknown length
            *      31960      32743:      contig of 784 bp in length
            *      gap of unknown length
            *      32744      33545:      contig of 802 bp in length
            *      gap of unknown length
            *      33546      34329:      contig of 784 bp in length
            *      gap of unknown length
            *      34330      35061:      contig of 732 bp in length
            *      gap of unknown length
            *      35062      35845:      contig of 784 bp in length
            *      gap of unknown length
            *      35846      36632:      contig of 787 bp in length
            *      gap of unknown length
            *      36633      37433:      contig of 801 bp in length
            *      gap of unknown length

```

\* 37434 38226: gap of unknown length  
\* contig of 793 bp in length  
\* 38227 38965: gap of unknown length  
\* contig of 739 bp in length  
\* 38966 39753: gap of unknown length  
\* contig of 788 bp in length  
\* 39754 40522: gap of unknown length  
\* contig of 769 bp in length  
\* 40523 41305: gap of unknown length  
\* contig of 783 bp in length  
\* 41306 42089: gap of unknown length  
\* contig of 784 bp in length  
\* 42090 42885: gap of unknown length  
\* contig of 796 bp in length  
\* 42886 43653: gap of unknown length  
\* contig of 788 bp in length  
\* 43654 44394: gap of unknown length  
\* contig of 741 bp in length  
\* 44395 45093: gap of unknown length  
\* contig of 699 bp in length  
\* 45094 45981: gap of unknown length  
\* contig of 788 bp in length  
\* 45982 46663: gap of unknown length  
\* contig of 782 bp in length  
\* 46664 47449: gap of unknown length  
\* contig of 786 bp in length  
\* 47450 48232: gap of unknown length  
\* contig of 783 bp in length  
\* 48233 49013: gap of unknown length  
\* contig of 781 bp in length  
\* 49014 49795: gap of unknown length  
\* contig of 782 bp in length  
\* 49796 50575: gap of unknown length  
\* contig of 780 bp in length  
\* 50576 51383: gap of unknown length  
\* contig of 808 bp in length  
\* 51384 52240: gap of unknown length  
\* contig of 857 bp in length  
\* 52241 53047: gap of unknown length  
\* contig of 807 bp in length  
\* 53048 53847: gap of unknown length  
\* contig of 800 bp in length  
\* 53848 54624: gap of unknown length  
\* contig of 777 bp in length  
\* 54625 55406: contig of 782 bp in length

Query Match 13.1%; Score 18; DB 43; Length 110000;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGACCTGGCTGAT 18

|||||

Db 85101 CCACCGACCTGGCTGAT 85084

RESULT 44

AC015845\_2/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC015845 Accession AC015845

Fragment Name	Begin	End
AC015845_0	1	110000
AC015845_1	100001	210000
AC015845_2	200001	310000
AC015845_3	300001	410000
AC015845_4	400001	510000
AC015845_5	500001	573550

Continuation (3 of 6) of AC015845 from base 200001 (AC015845 Homo sapiens chromosome 17

Query Match

Best Local Similarity 13.1%; Score 18; DB 43; Length 110000;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCACCGACCTGGCTGAT 18

|||||

Db 29319 CCACCGACCTGGCTGAT 29302

RESULT 45

AC013673

LOCUS

AC013673

DEFINITION

AC013673

ACCESSION

AC013673.1

VERSION

AC013673.1

KEYWORDS

HTG; HTGS\_PHASE0.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 121345)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 121345)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouhgalter,B.,

Brown,A., Castler,A., Colangelo,M., Collins,S., Collamore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,I., Doyle,M.,

Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieur,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html.

\* NOTE: This record contains 153 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads

\* and the order in which they appear is completely

\* arbitrary. Low-pass sequence sampling is useful for

\* identifying clones that may be gene-rich and allows

\* overlap relationships among clones to be deduced.

\* However, it should not be assumed that this clone

\* will be sequenced to completion. In the event that

\* the record is updated, the accession number will

\* be preserved.

\* 1

\* 824: contig of 824 bp in length

\* 825

\* gap of unknown length

\* 1622: contig of 798 bp in length

\* 1623

\* gap of unknown length

\* 2400: contig of 778 bp in length

\* 2401

\* gap of unknown length

\* 3201: contig of 801 bp in length

\* 3202

\* gap of unknown length

\* 3986: contig of 785 bp in length

\* 3987

\* gap of unknown length

\* 4779: contig of 793 bp in length

\* 4780

\* gap of unknown length

\* 5590: contig of 811 bp in length

\* 5591

\* gap of unknown length

\* 6387: contig of 797 bp in length

\* 6388

\* gap of unknown length

\* 7199: contig of 812 bp in length

\* 7200

\* gap of unknown length

\* 7948: contig of 749 bp in length

\* gap of unknown length

```
* 7949 8737: contig of 789 bp in length
* gap of unknown length
* 8738 9543: contig of 806 bp in length
* gap of unknown length
* 9544 10346: contig of 803 bp in length
* gap of unknown length
* 10347 11190: contig of 844 bp in length
* gap of unknown length
* 11191 12006: contig of 816 bp in length
* gap of unknown length
* 12007 12797: contig of 791 bp in length
* gap of unknown length
* 12798 13571: contig of 774 bp in length
* gap of unknown length
* 13572 14425: contig of 854 bp in length
* gap of unknown length
* 14426 15224: contig of 799 bp in length
* gap of unknown length
* 15225 16013: contig of 789 bp in length
* gap of unknown length
* 16014 16805: contig of 792 bp in length
* gap of unknown length
* 16806 17621: contig of 816 bp in length
* gap of unknown length
* 17622 18451: contig of 830 bp in length
* gap of unknown length
* 18452 19254: contig of 803 bp in length
* gap of unknown length
* 19255 20056: contig of 802 bp in length
* gap of unknown length
* 20057 20850: contig of 794 bp in length
* gap of unknown length
* 20851 21634: contig of 784 bp in length
* gap of unknown length
* 21635 22419: contig of 785 bp in length
* gap of unknown length
* 22420 23210: contig of 791 bp in length
* gap of unknown length
* 23211 23993: contig of 783 bp in length
* gap of unknown length
* 23994 24786: contig of 793 bp in length
* gap of unknown length
* 24787 25578: contig of 792 bp in length
* gap of unknown length
* 25579 26394: contig of 815 bp in length
* gap of unknown length
* 26394 27192: contig of 799 bp in length
* gap of unknown length
* 27193 28002: contig of 810 bp in length
* gap of unknown length
* 28003 28832: contig of 830 bp in length
* gap of unknown length
* 28833 29638: contig of 806 bp in length
* gap of unknown length
* 29639 30411: contig of 773 bp in length
* gap of unknown length
* 30412 31191: contig of 780 bp in length
* gap of unknown length
* 31192 32000: contig of 809 bp in length
* gap of unknown length
* 32001 32780: contig of 780 bp in length
* gap of unknown length
* 32781 33600: contig of 820 bp in length
* gap of unknown length
* 33601 34397: contig of 797 bp in length
* gap of unknown length
* 34398 35193: contig of 796 bp in length
* gap of unknown length
* 35194 35976: contig of 783 bp in length
* gap of unknown length
* 35977 36772: contig of 796 bp in length
* gap of unknown length
* 36773 37564: contig of 792 bp in length
```

```
* 37565 38346: contig of 782 bp in length
* gap of unknown length
* 38347 39123: contig of 777 bp in length
* gap of unknown length
* 39124 39921: contig of 798 bp in length
* gap of unknown length
* 39922 40723: contig of 802 bp in length
* gap of unknown length
* 40724 41523: contig of 800 bp in length
* gap of unknown length
* 41524 42358: contig of 835 bp in length
* gap of unknown length
* 42359 43174: contig of 816 bp in length
* gap of unknown length
* 43175 43989: contig of 815 bp in length
* gap of unknown length
* 43990 44779: contig of 790 bp in length
* gap of unknown length
* 44780 45569: contig of 790 bp in length
* gap of unknown length
* 45570 46371: contig of 802 bp in length
* gap of unknown length
* 46372 47169: contig of 798 bp in length
* gap of unknown length
* 47170 47957: contig of 788 bp in length
* gap of unknown length
* 47958 48772: contig of 815 bp in length
* gap of unknown length
* 48773 49548: contig of 776 bp in length
* gap of unknown length
* 49549 50338: contig of 790 bp in length
* gap of unknown length
* 50339 51143: contig of 805 bp in length
* gap of unknown length
* 51144 51943: contig of 800 bp in length
* gap of unknown length
* 51944 52666: contig of 723 bp in length
* gap of unknown length
* 52667 53437: contig of 771 bp in length
* gap of unknown length
* 53438 54240: contig of 803 bp in length
* gap of unknown length
* 54241 55042: contig of 802 bp in length
* gap of unknown length
* 55043 55827: contig of 785 bp in length
* gap of unknown length
* 55828 56632: contig of 805 bp in length
* gap of unknown length
* 56633 57449: contig of 817 bp in length
* gap of unknown length
* 57450 58278: contig of 829 bp in length
* gap of unknown length
* 58279 59053: contig of 775 bp in length
* gap of unknown length
* 59054 59838: contig of 785 bp in length
* gap of unknown length
* 59839 60636: contig of 798 bp in length
* gap of unknown length
* 60637 61415: contig of 779 bp in length
* gap of unknown length
* 61416 62222: contig of 807 bp in length
```

Query Match 13.1% Score 18; DB 43; Length 121345;  
Best Local Similarity 100.0%; Pred. No. 5.8;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 CTCCCAAGCTAAACAAT 100  
|||||  
DB 10859 CTCCCAAGCTAAACAAT 10876

Search completed: April 6, 2000, 21:20:42  
Job time: 50627 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 5, 2000, 11:35:47 ; Search time 6420.76 Seconds  
(without alignments)  
161.123 Million cell updates/sec

Title: US-09-090-672B-12  
Perfect score: 274  
Sequence: 1 CGTTACAGATCTCTTGGC.....GAACATGAAAAATGAAAAAT 274

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues  
Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est30:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_gss1:\*  
80: gb\_gss2:\*  
81: gb\_gss3:\*  
82: gb\_gss4:\*  
83: em\_gss1:\*  
84: em\_gss2:\*  
85: em\_gss3:\*  
86: em\_gss4:\*  
87: gb\_gss5:\*  
88: gb\_gss6:\*  
89: gb\_gss7:\*  
90: gb\_gss8:\*  
91: gb\_gss9:\*  
92: em\_gss5:\*  
93: em\_gss6:\*  
94: em\_gss7:\*  
95: em\_gss8:\*  
96: em\_gss9:\*  
97: em\_gss10:\*  
98: em\_gss11:\*  
99: gb\_gss10:\*  
100: gb\_gss11:\*  
101: em\_gss12:\*  
102: gb\_gss12:\*  
103: gb\_gss13:\*  
104: gb\_gss14:\*  
105: gb\_gss15:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	198.6	72.5	405	21	T98890	T98890 ye65f09.r1
2	138	50.4	493	23	H18435	H18435 ym43b05.s1

```

c 3 127.4 46.5 302 20 F03205
4 117.8 43.0 390 39 AA878923
5 113 41.2 318 30 AA197894
6 113 41.2 459 39 AA873942
7 111.4 40.7 731 74 AA211128
8 88.4 32.3 391 33 AA432984
9 81.8 29.9 387 22 H09173
10 74.6 27.2 274 32 AA334229
11 64.6 23.6 559 34 AA488906
12 64.2 23.4 259 20 T10350
13 63.6 23.2 596 104 AQ587717
14 62.8 22.9 449 43 A1196903
15 53.8 19.6 614 79 FR0008178
16 47.6 17.4 1203 80 CNS015WU
17 45.4 16.6 354 22 R45421
18 41.8 15.3 1201 80 CNS0167M
19 41.6 15.2 624 105 AQ620120
20 41.2 15.0 242 45 A1394828
21 41 15.0 483 49 A1632118
22 40.8 14.9 583 104 AQ570963
23 40.6 14.8 437 38 AA755877
24 40.4 14.7 415 35 AA550066
25 40 14.6 423 88 AQ836402
26 40 14.6 488 91 AQ129266
27 39.8 14.5 1101 80 CNS016DT
28 39.6 14.5 1101 79 CNS008X3
29 39.4 14.4 467 80 CNS018G8
30 39.4 14.4 1101 80 CNS0176D
31 39 14.2 1101 80 CNS0142R
32 39 14.2 1204 80 CNS016E2
33 38.8 14.2 346 60 A1783824
34 38.4 14.0 561 91 AQ139258
35 38.4 14.0 618 100 AQ326393
36 38.4 14.0 654 99 AQ259988
37 38.4 14.0 928 79 CNS00DKY
38 38.4 14.0 1101 79 CNS003B6
39 37.8 13.8 383 45 A1394910
40 37.6 13.7 259 73 A1373585
41 37.4 13.6 455 45 A1395037
42 37.2 13.6 229 45 A1394956
43 37.2 13.5 778 79 CNS010L9
44 37 13.5 299 61 A1860086
45 37 13.5 441 61 A1811918

```

## ALIGNMENTS

```

RESULT 1
T98890 405 bp mRNA EST 31-MAR-1995
LOCUS ye65f09.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:122633 5', mRNA sequence.

ACCESSION T98890
VERSION T98890.1 GI:748627
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 405)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

```

FEATURES  
Source

Email: est@watson.wustl.edu  
Insert Size: 1065  
High quality sequence stops: 279 Source: IMAGE Consortium, LLNL This  
clone is available royalty-free through LLNL; contact the IMAGE  
Consortium (info@image.llnl.gov) for further information.  
Insert length: 1065 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 279.

Location/Qualifiers

1..405  
/organism="Homo sapiens"  
/db\_xref="GDB:475178"  
/db\_xref="taxon:9606"  
/clone="IMAGE:122633"  
/clone.lib="Soares fetal liver spleen INFLS"  
/sex="Male"  
/dev\_stage="20 week-post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;  
1st strand cDNA was primed with a Pac I - oligo(dT) primer  
[5', AACTGGAAGAATAAATAAGATCTTTTTTTTTTTTTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 116 a 68 c 133 t  
ORIGIN

Query Match 72.5%; Score 198.6; DB 21; Length 405;  
Best Local Similarity 98.7%; Pred. No. 2.1e-36;  
Matches 220; Conservative 1; Mismatches 0; Indels 2; Gaps 2;  
QY 31 GAACATAAAGGGATCGGTCCTATATACAAATACCAACATGTAATAAATAGATTCT 90  
Db 2 GAACATAAAGGGATCGGTCCTATATACAAATACCAACATGTAATAAATAGATTCT 61  
QY 91 GTGTTCCTGCTTATAGACCATGTTTGTAGTAGGTAAGAGAAACCTCTATATCTGAA 150  
Db 62 GTGTTCCTGCTTATAGACCATGTTTGTAGTAGGTAAGAGAAACCTCTATATCTGAA 121  
QY 151 ACAGCTTAACATTTACAAAATTTAGTTTCTTTTGTAGAGTCTTATCCGTGA-GCTAT 209  
Db 122 ACAGCTTAACATTTACAAAATTTAGTTTCTTTTGTAGAGTCTTATCCGTGCTAT 181  
QY 210 ATACAGTTCATGCTGATTGA-GCATTGTTTCACGAGTAAAG 251  
Db 182 ATACAGTTCATGCTGATTGAGCATTGTTTCACGAGTAAAG 224

RESULT 2

LOCUS H18435/c 493 bp mRNA EST 29-JUN-1995  
DEFINITION Ym43B05.sl Soares infant brain INIB Homo sapiens cDNA clone  
IMAGE:50988 3', mRNA sequence.  
ACCESSION H18435  
VERSION H18435.1 GI:884675  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 493)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
Wilson,R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT On May 18, 1995 this sequence version replaced gi:811454.



Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1071  
High quality sequence stops: 327  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL ; con  
IMAGE Consortium (infoimage.lnl.gov) for further info  
Insert Length: 1071 Std Error: 0.00  
seq primer: Promega -21mi3  
High quality sequence stop: 327.

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital status</i>	3. <i>Marital status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Income</i>	5. <i>Income</i>
6. <i>Occupation</i>	6. <i>Occupation</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political affiliation</i>	8. <i>Political affiliation</i>
9. <i>Health status</i>	9. <i>Health status</i>
10. <i>Family size</i>	10. <i>Family size</i>
11. <i>Home ownership</i>	11. <i>Home ownership</i>
12. <i>Travel frequency</i>	12. <i>Travel frequency</i>
13. <i>Spending habits</i>	13. <i>Spending habits</i>
14. <i>Interests and hobbies</i>	14. <i>Interests and hobbies</i>
15. <i>Volunteer work</i>	15. <i>Volunteer work</i>
16. <i>Charitable contributions</i>	16. <i>Charitable contributions</i>
17. <i>Political participation</i>	17. <i>Political participation</i>
18. <i>Community involvement</i>	18. <i>Community involvement</i>
19. <i>Environmental awareness</i>	19. <i>Environmental awareness</i>
20. <i>Technological usage</i>	20. <i>Technological usage</i>
21. <i>Healthcare utilization</i>	21. <i>Healthcare utilization</i>
22. <i>Insurance coverage</i>	22. <i>Insurance coverage</i>
23. <i>Retirement planning</i>	23. <i>Retirement planning</i>
24. <i>Investment preferences</i>	24. <i>Investment preferences</i>
25. <i>Real estate interests</i>	25. <i>Real estate interests</i>
26. <i>Art and cultural consumption</i>	26. <i>Art and cultural consumption</i>
27. <i>Food and dining preferences</i>	27. <i>Food and dining preferences</i>
28. <i>Exercise and fitness routines</i>	28. <i>Exercise and fitness routines</i>
29. <i>Travel preferences</i>	29. <i>Travel preferences</i>
30. <i>Charitable giving patterns</i>	30. <i>Charitable giving patterns</i>
31. <i>Political engagement</i>	31. <i>Political engagement</i>
32. <i>Community service hours</i>	32. <i>Community service hours</i>
33. <i>Environmental activism</i>	33. <i>Environmental activism</i>
34. <i>Digital footprint</i>	34. <i>Digital footprint</i>
35. <i>Healthcare choices</i>	35. <i>Healthcare choices</i>
36. <i>Insurance policy details</i>	36. <i>Insurance policy details</i>
37. <i>Retirement account types</i>	37. <i>Retirement account types</i>
38. <i>Investment portfolio composition</i>	38. <i>Investment portfolio composition</i>
39. <i>Real estate portfolio</i>	39. <i>Real estate portfolio</i>
40. <i>Art collection details</i>	40. <i>Art collection details</i>
41. <i>Dining frequency and preferences</i>	41. <i>Dining frequency and preferences</i>
42. <i>Exercise frequency and intensity</i>	42. <i>Exercise frequency and intensity</i>
43. <i>Travel frequency and destinations</i>	43. <i>Travel frequency and destinations</i>
44. <i>Charitable giving frequency and amounts</i>	44. <i>Charitable giving frequency and amounts</i>
45. <i>Political participation frequency</i>	45. <i>Political participation frequency</i>
46. <i>Community involvement frequency</i>	46. <i>Community involvement frequency</i>
47. <i>Environmental activism frequency</i>	47. <i>Environmental activism frequency</i>
48. <i>Digital footprint details</i>	48. <i>Digital footprint details</i>
49. <i>Healthcare utilization frequency</i>	49. <i>Healthcare utilization frequency</i>
50. <i>Insurance coverage details</i>	50. <i>Insurance coverage details</i>
51. <i>Retirement planning frequency</i>	51. <i>Retirement planning frequency</i>
52. <i>Investment preferences frequency</i>	52. <i>Investment preferences frequency</i>
53. <i>Real estate interests frequency</i>	53. <i>Real estate interests frequency</i>
54. <i>Art and cultural consumption frequency</i>	54. <i>Art and cultural consumption frequency</i>
55. <i>Food and dining preferences frequency</i>	55. <i>Food and dining preferences frequency</i>
56. <i>Exercise and fitness routines frequency</i>	56. <i>Exercise and fitness routines frequency</i>
57. <i>Travel preferences frequency</i>	57. <i>Travel preferences frequency</i>
58. <i>Charitable giving patterns frequency</i>	58. <i>Charitable giving patterns frequency</i>
59. <i>Political engagement frequency</i>	59. <i>Political engagement frequency</i>
60. <i>Community service hours frequency</i>	60. <i>Community service hours frequency</i>
61. <i>Environmental activism frequency</i>	61. <i>Environmental activism frequency</i>
62. <i>Digital footprint frequency</i>	62. <i>Digital footprint frequency</i>
63. <i>Healthcare choices frequency</i>	63. <i>Healthcare choices frequency</i>
64. <i>Insurance policy details frequency</i>	64. <i>Insurance policy details frequency</i>
65. <i>Retirement account types frequency</i>	65. <i>Retirement account types frequency</i>
66. <i>Investment portfolio composition frequency</i>	66. <i>Investment portfolio composition frequency</i>
67. <i>Real estate portfolio frequency</i>	67. <i>Real estate portfolio frequency</i>
68. <i>Art collection details frequency</i>	68. <i>Art collection details frequency</i>
69. <i>Dining frequency and preferences frequency</i>	69. <i>Dining frequency and preferences frequency</i>
70. <i>Exercise frequency and intensity frequency</i>	70. <i>Exercise frequency and intensity frequency</i>
71. <i>Travel frequency and destinations frequency</i>	71. <i>Travel frequency and destinations frequency</i>
72. <i>Charitable giving frequency and amounts frequency</i>	72. <i>Charitable giving frequency and amounts frequency</i>
73. <i>Political participation frequency</i>	73. <i>Political participation frequency</i>
74. <i>Community involvement frequency</i>	74. <i>Community involvement frequency</i>
75. <i>Environmental activism frequency</i>	75. <i>Environmental activism frequency</i>
76. <i>Digital footprint frequency</i>	76. <i>Digital footprint frequency</i>
77. <i>Healthcare choices frequency</i>	77. <i>Healthcare choices frequency</i>
78. <i>Insurance policy details frequency</i>	78. <i>Insurance policy details frequency</i>
79. <i>Retirement planning frequency</i>	79. <i>Retirement planning frequency</i>
80. <i>Investment preferences frequency</i>	80. <i>Investment preferences frequency</i>
81. <i>Real estate interests frequency</i>	81. <i>Real estate interests frequency</i>
82. <i>Art and cultural consumption frequency</i>	82. <i>Art and cultural consumption frequency</i>
83. <i>Food and dining preferences frequency</i>	83. <i>Food and dining preferences frequency</i>
84. <i>Exercise and fitness routines frequency</i>	84. <i>Exercise and fitness routines frequency</i>
85. <i>Travel preferences frequency</i>	85. <i>Travel preferences frequency</i>
86. <i>Charitable giving patterns frequency</i>	86. <i>Charitable giving patterns frequency</i>
87. <i>Political engagement frequency</i>	87. <i>Political engagement frequency</i>
88. <i>Community service hours frequency</i>	88. <i>Community service hours frequency</i>
89. <i>Environmental activism frequency</i>	89. <i>Environmental activism frequency</i>
90. <i>Digital footprint frequency</i>	90. <i>Digital footprint frequency</i>
91. <i>Healthcare choices frequency</i>	91. <i>Healthcare choices frequency</i>
92. <i>Insurance policy details frequency</i>	92. <i>Insurance policy details frequency</i>
93. <i>Retirement planning frequency</i>	93. <i>Retirement planning frequency</i>
94. <i>Investment preferences frequency</i>	94. <i>Investment preferences frequency</i>
95. <i>Real estate interests frequency</i>	95. <i>Real estate interests frequency</i>
96. <i>Art and cultural consumption frequency</i>	96. <i>Art and cultural consumption frequency</i>
97. <i>Food and dining preferences frequency</i>	97. <i>Food and dining preferences frequency</i>
98. <i></i>	

```

/organism="Homo sapiens"
/db_xref="GDB:423736"
/db_xref="taxon:9606"
/clone_lib="IMAGE:50988"
/clone_lib="Soares infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: LaMid BA; Site: 1; Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I...Oligo(dT) primer [5';
ACTGGAGAAATTCGGCGGCAGCAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the LaMid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
86 C 106 G 148 T 7 others

```

BASE COUNT	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

```
Query Match      50.4%; Score 138; DB 23; Length 493;
Best Local Similarity 77.3%; Pred. No. 1.7e-22;
Matches 218: Conservative 0; Mismatches 54; Indels 10; Gaps 4;
```

Qy	3	TTTACAGATTCTCTTGGCGCTGGCGTGGAGACTACAAAGGATCGGTG-CCTATATACAA	61
Db	466	TTAAACNAATCCCCCGCGAGACGGNAGAACCTACAANGGAACGGTGCCTTATATACAA	407
Qy	62	ATACCAAACTTGATAATAATCTAGATCTCTGTCTGCTTATAGACCATGTTTGT- ---	118
Db	406	ATACCCAAACTGATAATAAACCTAGATCTCTGTCTCCCGCCTAATAGACCCATGTTTGT	347
Qy	118	-AGTAGGTAGAGGAAACATTCC- ---TATATTCTGAAACAGCCTACATTTTACAAAATT	173
Db	346	AAGTAGGTAGAGGAAACATTCCCTATANTCCTGAAACAGCCCTAAACATTTACAAAATT	287
Qy	174	TTAGTTTTCTTTTTTAGAGCTTTATCTCTGTAG-CATATAAACAGTTCATGCTGATTTAG	232
Db	286	TTAGTTTTCTTTTTTAGAGCTTTATCTCTGTAGCCTATATAACAGTTCATGCTGATTTAG	227
Qy	233	CATTGTGTCAGGAGTAAGCTGGAACATATGAAAATTGAAAAT	274
Db	226	CATTGTGTCAGGAGTAAGCTGGAACATATGAAAATTGAAAAT	185

RESULT	3
F03205/0	

LOCUS	F03205	302 bp	mRNA	EST	02-FEB-1995			
DEFINITION	HSC1bF072 normalised infant brain cDNA Homo sapiens CDNA clone c-1p07 3', mRNA sequence.							
ACCESSION	F03205							
VERSION	F03205.1	GI:646762						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;							

REFERENCE  
AUTHORS

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
 Insert Size: 1071

Insert size: 1071  
High quality sequence stops: 327

Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1071 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 327

FEATURES SOURCE

Email: [genexpress@genethon.fr](mailto:genexpress@genethon.fr)  
Single read, removed at sequence 5'end  
Genexpress\_library\_id: C; Genexpress\_sequence\_id: alc-  
Seq primer: (-21)M13 universal.

FEATURES  
SOURCE

```

1:..302
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-lpf07"
/clone_fib="normalized infant brain cdna"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"

```

Age=Stage=3 months; Sex=Male; Vector: lsfmid BA; Site=1: HindIII; Note=Organ: brain; dev\_stage=3 months old; Site=2: NotI; sex=Female; dev\_stage=3 months old; Isolate=muscular atrophy patient; tissue\_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lsfmid BA vector. Clone library from B Soares, Psychiatry Dept, Columbia University, USA. Normalization method:

BASE COUNT	111 a	54 c	52 g	84 t	1 others
ORIGIN					

Query Match	46.5%	Score 127.4;	DB 20;	Length 302;
Best Local Similarity	99.2%	Pred. No. 5.2e-20;		
Matches 128: Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY 146 CTGAACAGCCTAACATTTTACAAATTTTAGTTTCTTTTATAGACTCTTATCCTGTAG 205

Db 302 CTGAACAGCCTAACATTTACAAATTTTAGTTTCTTTTTTAGAGTCTTATCCTGTAG 243

QY 206 CTATATAACAGTTCATGTCCTGATTTAGCATTTGTTCCAGAGTAAGCTGGAACATATGAAA 265

Db 242 CTATATAACAGTTCATGTCGTGATTTAGCATTTGTCACGAGTAAAGCTGGACAAATGAAA 183

QY 266 ATTGAAT 274

Db 182 ATTGAAAT 174

## RESULT 4

AA878923	AA878923	390 bp	mRNA	EST	19-MAY-1998
LOCUS	AA878923				
DEFINITION	oj25d11.s1 NCI_CGAP_Kid5 Homo sapiens CDNA clone IMAGE:149320 similar to TR:014432 O14432 TUP1.1, mRNA sequence.				

ACCESSION AA878923  
VERSION AA878923.1 GI:2987888  
KEYWORDS EST.  
SOURCE human.

**SOURCE:** Human.  
**ORGANISM:** Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE:** 1 (bases 1 to 390)  
**AUTHORS:** NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE:** National Cancer Institute, Cancer Genome Anatomy Project (CGAP)

**JOURNAL  
COMMENT**

**Tumor Gene Index**  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2152443.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1846    Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 365.

[illegible]

I. .390  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1493205"  
/clone\_lib="NCI\_CGAP\_Kid5"  
/tissue\_type="2 pooled tumors (clear cell type)"  
/lab\_host="DH10B"  
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'- AACTGGAGAATTCGGCGGCGCAATATTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7t3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
86 c 87 g 107 t

```

BASE COUNT      110 a      86 c      87 g      107 t
ORIGIN
      Constructed by GenCo Soares and M. Fatima Bomardo.

Query Match      43.0%; Score 117.8; DB 39; Length 390;
Best Local Similarity 96.7%; Pred. No. 8.1e-18;
Matches 119; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 CGTTTACAGATTCCTTCGGCGGTGGCGGTGGAACCTCAAAAGGGATCGGTCCTATATCAC 60
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      30 CGTTTACAGACTCTCTCGCGCAGCGGTGGAACCTCAAAAGGGATCGGTCCTATATCAC 89
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY      61 AATACCAAACTTGATAATAATCTAGATTCGTGTYCTCTTATAGACCATGTTTGTAGT 120
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      90 AATACCAAACTTGATAATAATCTAGATTCGTGTCCTGCTTATAGACCATGTTTGTAGT 149
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY      121 AGG 123
      |||
Db      150 AGG 152

```

RESULT	5
AAI97894	
LOCUS	318 bp mRNA EST 12-MAR-1997
DEFINITION	m05e02.r1 GuayWoodford Beiler mouse kidney day 0 Mus musculus cdNA clone IMAGE:654170 5', mRNA sequence.
ACCESSION	AAI97894
VERSION	AAI97894.1 GI:1793528
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 318)
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston, R.  
The WashU-HMMI Mouse EST Project  
Unpublished (1996)  
On Sep 12, 1996 this sequence version replaced g1:1394222.  
Contact: Marra M/Mouse EST Project  
WashU-HMMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:400018

seq primer: -28013 rev1 ET from Anne's main  
High quality sequence stop: 297.

FEATURES	SOURCE
1. <b>Feature 1</b>	Source 1
2. <b>Feature 2</b>	Source 2
3. <b>Feature 3</b>	Source 3
4. <b>Feature 4</b>	Source 4
5. <b>Feature 5</b>	Source 5
6. <b>Feature 6</b>	Source 6
7. <b>Feature 7</b>	Source 7
8. <b>Feature 8</b>	Source 8
9. <b>Feature 9</b>	Source 9
10. <b>Feature 10</b>	Source 10

```

1. 310
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="654170"
/clone_lib="GuayWoodford Beier mouse kidney day 0"
/tissue_type="kidney"
/dev_stage="newborn (day 0)"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: pBluescript SK-; Site: EcoRI; Site 2: XhoI; Cloned unidirectionally. Pri- Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR -5' adaptor sequence: 5' GAAATCGGCACGAG 3', -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3' Library p Lisa Guay-Woodford."

```

BASE COUNT	89 a	74 c	73 g	82 t
------------	------	------	------	------

Query Match	41.2%;	Score 113;	DB 30;	Length 318;
Best Local Similarity	94.3%;	Pred. No. 1.1e-16;		
Matches 116; Conservative	1;	Mismatches 6;	Indels	

QY	1	CGTTTACAGATTCCTCTGCGGCTGGCGGCTGAACTACAAGGGATCGGTGCTATATCAC	60
Db	112	CGTTTACAGACTCTCCCTGCGCAGCGGTGAACTACAAGGGATCGGTGCTATATCAC	171
QY	61	AATACCAAACCTTGATAATAATCTAGATCTCTGTYTCTGCTTATAGACCATTGTTGTAGT	120
Db	172	AATACCAAACCTTGACAATAATCTAGATCTCTGCTCTCTACTTATAGATCATGTTGTAGT	231
QY	121	AGG	123
Db	232	AGG	234

## RESULT

AA873942

LOCUS	AA873942	459 bp	mrna	EST	19-MAR-1998
DEFINITION	vw86a03.r1 Stratiogene mouse skin (4937313) Mus musculus cDNA clone IMAGE:1261804 5' similar to TR:O14432 O14432 TUF1. ;, mRNA sequence.				

sequence.  
AA873942

ACCESSION  
VERSION  
AA873942.1  
AA873942.1  
GI:2978631

**KEYWORDS**  
EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 459)

**AUTHORS**  
Marra, M., Hillier, I., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gelsiel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project

**TITLE**



Waterston, R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
On May 9, 1995 this sequence version replaced gi:804216.  
Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:493399

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 358.

#### FEATURES

source  
Location/Qualifiers  
1..331  
/organism="Mus musculus"  
/strain="B6D2 F1/J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:833183"  
/clone\_lib="Knowles Solter mouse 2 cell"  
/tissue\_type="embryo"  
/dev\_stage="2-cell"  
/lab\_host="DH108"

/note="Organ: embryo; Vector: pBluescribe (modified);  
Site1: MluI; Site2: SalI; Cloned unidirectionally from  
mRNA prepared from 13 500 2-cell stage embryos. Primer:  
SalI(UT): 5'-CGGTCAGCGTCGACGGTCTTTTCTTTT-3'. cDNAs  
were cloned into the MluI/SalI sites of a modified  
pBluescribe vector using commercial linkers (NEB).  
Average insert size: 1.2 kb."

BASE COUNT 118 a 83 c 96 g 94 t

#### ORIGIN

Query Match 32.3%; Score 88.4; DB 33; Length 391;  
Best Local Similarity 84.5%; Pred. No. 4.7e-11;  
Matches 98; Conservative 1; Mismatches 17; Indels 0; Gaps 0;  
QY 8 AGATTCTCTGGCGTGGCTGAACACAAAGGGATCGGTGCTATATACAAATACCA 67  
|||||  
Db 102 AGATCCGGGGCGCGGTGGTGAACACAAAGGGATCGGTGCTATATACCAATACCA 161  
QY 68 AACTTGATAATCTAGATTCTGTGTCTGTCTATAGACCATGTTTGTAGTAGG 123  
|||||  
Db 162 AACTGGACAAATCTAGATTCTGTGTCTCTACTTATAGATCATGTTTGTAGTAGG 217

#### RESULT 9

H09173/c 387 bp mRNA EST 23-JUN-1995  
LOCUS H09173  
DEFINITION Y198h02.s1 Soares infant brain lNIB Homo sapiens cDNA clone  
IMAGE:46519 3', mRNA sequence.  
ACCESSION H09173  
VERSION H09173.1 GI:873395  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 387)

REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

THE WASHU-MERCK EST PROJECT

Unpublished (1995)

On May 5, 1995 this sequence version replaced gi:797954.

Contact: Wilison RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 1081  
High quality sequence stops: 314

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 1081 Std Error: 0.00

Seq primer: Promega -2ml3

High quality sequence stop: 314.

#### FEATURES

source  
Location/Qualifiers  
1..387  
/organism="Homo sapiens"  
/db\_xref="GDB:419060"  
/db\_xref="taxon:9606"  
/clone\_lib="Soares infant brain lNIB"  
/sex="female"

/dev\_stage="73 days post natal"

/lab\_host="DH108 (ampicillin resistant)"

/note="Organ: whole brain; Vector: Lfamid BA; Site1: Not  
I; Site2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5'

AACTGGAAGAAATTCGCCGCCGAGGAATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the Lfamid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 132 a 67 c 72 g 112 t 4 Others

#### ORIGIN

Query Match 29.9%; Score 81.8; DB 22; Length 387;

Best Local Similarity 85.7%; Pred. No. 1.5e-09;

Matches 150; Conservative 0; Mismatches 17; Indels 8; Gaps 5;

QY 108 CCATGTTTGTAGTAGTA-AGAGGAAACTTC--TATATTCTGAACAGCCTACATT 163

|||||

Db 359 CCATGTTTGTAGTAGTANAGAGAAACTTCCTATATATCTGGAACAGCCTACATT 300

|||||

QY 164 TTACAAAATTTTAGTTTCTTTTITTAGAGT--CTTATCTGTAG-CTATATACAGTTCA 220

|||||

Db 299 TTACAAAATTTTAGTTTCTTTTITTAGAGTCTTATCTGTAGCCTATATACAGTTCA 240

|||||

QY 221 TGT-CTGATTAGCATTTGTCACGAGTAAGCTGGAACTATGAAATTTGAAAT 274

|||||

Db 239 TGTCTGATTAGCATTTGTCACGAGTAAGCTGGAACTATGAAATTTGAAAT 185

#### RESULT 10

AA334229 274 bp mRNA EST 21-APR-1997  
LOCUS AA334229  
DEFINITION EST38412 Embryo, 9 week Homo sapiens cDNA 5' end, mRNA sequence.  
ACCESSION AA334229  
VERSION AA334229.1 GI:1986472  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 274)

REFERENCE

AUTHORS

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man- Wal, C.,  
Clayton, R.A., Cline, R.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,  
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrle, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
12140200

**TITLE** On Dec 3, 1996 this sequence version replaced gi:1119186.

**JOURNAL** Other ESTs: TH0126123

**MEDLINE** Contact: Kerlavage, AR

**COMMENT** Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tei: 3018699056

Fax: 3018699423

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tadb/hgi/hgi.html>)

Seq primer: M13 Reverse.

**FEATURES** Location/Qualifiers

1..274

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):135902"

/db\_xref="taxon:9606"

/clone\_lib="Embryo, 9 week"

/dev\_stage="embryo, 9 wks"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI

87 a 51 c 53 g 81 t 2 others

**BASE COUNT** 87 a 51 c 53 g 81 t 2 others

**ORIGIN**

Query Match 27.2%; Score 74.6; DB 32; Length 274;  
Best Local Similarity 98.7%; Pred. No. 7.4e-08;  
Matches 74; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 49 TGCCTATATACAAATACCAACTTGATAATATCTAGATTCTGTGTCCTATATAGAC 108

|||||

Db 1 TGCCTATATACAAATACCAACTTGATAATATCTAGATTCTGTGTCCTATATAGAC 50

QY 109 CATGTTGTAGTAGG 123

|||||

Db 61 CATGTTGTAGTAGG 75

**RESULT 11**  
AA488906/c  
LOCUS AA488906 559 bp mRNA EST 15-AUG-1997  
DEFINITION aa55a02.r1 NCI-CGAP-GCB1 Homo sapiens cDNA clone IMAGE:824810 5' similar to TR:G507003 G507003 BETA TRANSDUCIN-LIKE PROTEIN. ;, mRNA sequence.

**ACCESSION** AA488906.1 GI:2218508

**VERSION** AA488906.1

**KEYWORDS** EST.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 559)

**REFERENCE** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

**AUTHORS** National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

**TITLE** Tumor Gene Index

**JOURNAL** Unpublished (1997)

**COMMENT** On Sep 12, 1996 this sequence version replaced gi:1397871.

Contact: Robert Strausberg, Ph.D.

Tei: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
cDNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 EF from Amersham

High quality sequence stop: 446.

**FEATURES** Location/Qualifiers

1..559

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:824810"

/clone\_lib="NCI-CGAP-GCB1"

/class\_type="germinal center B cell"

/lab\_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA

was prepared from human tonsillar cells enriched for

germinal center B cells by flow sorting (CD20+, IgD+),

provided by Dr. Louis M. Staudt (NCI). Dr. David Allman

(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was

primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTT-

3']. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

158 a 139 c 100 g 162 t

**BASE COUNT** 158 a 139 c 100 g 162 t

**ORIGIN**

Query Match 23.6%; Score 64.6; DB 34; Length 559;  
Best Local Similarity 98.5%; Pred. No. 1.3e-05;  
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 59 ACATACCAAACTTCAATATCTAGATTCTGTCTCTATAGACCATGTTTGA 118

|||||

Db 559 ACAATACCAAACTTCAATATCTAGATTCTGTCTCTATAGACCATGTTTGA 500

QY 119 GTAGG 123

|||||

Db 499 GTAGG 495

**RESULT 12**  
T10350/c  
LOCUS T10350 259 bp mRNA EST 13-APR-1994  
DEFINITION seq1091 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone b4HB3MA-COT8-HAP-Ft287 3', mRNA sequence.

**ACCESSION** T10350

**VERSION** T10350.1 GI:471699

**KEYWORDS** EST.

**SOURCE** human.

**ORGANISM** Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 259)

**REFERENCE** Soares, M.B., Bonaldo, M.F., Jelenc, P., Su, L., Lawton, L. and

Efstathiadis, A.

**AUTHORS** Construction and characterization of a normalized cDNA library

**TITLE** Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)

**JOURNAL** 95023884

**COMMENT** Contact: Bento Soares

Columbia University

Department of Psychiatry, 722 W 168th Street, Unit #41, New York, NY

10032



BASE COUNT 135 a 82 c 105 g 127 t  
ORIGIN

Query Match 22.9%; Score 62.8; DB 43; Length 449;  
Best Local Similarity 94.1%; Pred. No. 3.5e-05;  
Matches 64; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 56 ATCAATACCAAACTTGATAAATCTAGATTCTGTGTCTGTATAGACCATGTTT 115  
|||||  
Db 1 ATCAATACCAAACTTGACAATAATCTAGATTCTGTGTCTTACTTATAGATCAATGTTT 60  
|||||

Qy 116 GTAGTAGG 123  
|||||  
Db 61 GTAGTAGG 68  
|||||

## RESULT 15

FR0008178  
LOCUS FR0008178 614 bp DNA GSS 02-MAR-1997  
DEFINITION F. rubripes GSS sequence, clone 188G03ad4, genomic survey sequence.  
ACCESSION 291988  
VERSION 291988.1 GI:1869202  
KEYWORDS GSS; genome survey sequence.  
SOURCE Fugu rubripes.  
ORGANISM Fugu rubripes.  
Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;  
Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;  
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Fugu.  
1 (bases 1 to 614)  
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrانيا, Y.,  
Williams, G. and Brenner, S.  
Direct Submission  
Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk  
Vector: ml3mp18  
V-type: phage  
PRIMER: M13  
DESCR:  
One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.

## FEATURES

source Location/Qualifiers  
1..614  
/organism="Fugu rubripes"  
/db\_xref="taxon:31033"  
/clone\_lib="cosmid 188G03"  
/clone="188G03ad4"

BASE COUNT 112 a 159 c 159 g 145 t 39 others  
ORIGIN

Query Match 19.6%; Score 53.8; DB 79; Length 614;  
Best Local Similarity 61.2%; Pred. No. 0.004;  
Matches 85; Conservative 1; Mismatches 53; Indels 0; Gaps 0;

Qy 1 CGTTTACAGATCTCTTCGGCTGGCGGCTGGAACCTACAAAGGATCGGTGCCCTATATCAC 60  
|||  
Db 99 CGCCTGACAGCTCTGCTCAAAACGGCGGTGGAGCTGCAGCGGGAACGTTGCTCTACCCAC 158  
|||  
Qy 61 AATACCAAACTTGATAATATCTAGATTCTGTGTCTGTCTATAGACCATGTTTGTAGT 120  
|||  
Db 159 AACACCAAGAGGAGGAGGAGGAGTGGACTGGTGGCCCTCTCTCGACCACTGCAC 218  
|||  
Qy 121 AGGTAAGAGGAAACTTCC 139  
|||  
Db 219 CGGTAAGGCAACCGGTCC 237  
|||

Search completed: April 5, 2000, 11:35:51  
Job time: 22828 sec

